

An Efficient Derivative-Free Milstein Scheme for Stochastic Partial Differential Equations with Commutative Noise

Claudine Leonhard* and Andreas Rößler†

Institute of Mathematics, Universität zu Lübeck,
Ratzeburger Allee 160, 23562 Lübeck, Germany

Abstract

We propose a derivative-free Milstein scheme for stochastic partial differential equations with trace class noise which fulfill a certain commutativity condition. The same theoretical order of convergence with respect to the spatial and time discretizations as for the Milstein scheme is obtained whereas the computational cost is, in general, considerably lower. As the main result, we show that the effective order of convergence of the proposed derivative-free Milstein scheme is significantly higher than the effective order of convergence of the original Milstein scheme if errors versus computational costs are considered. In addition, the derivative-free Milstein scheme is efficiently applicable to general semilinear stochastic partial differential equations which do not have to be multiplicative in the Q -Wiener process. Finally we prove the convergence of the proposed scheme. This version of the paper presents work in progress.

1 Introduction

Stochastic partial differential equations (SPDEs) are a powerful tool to model various phenomena from biology to finance. They can be employed, for example, to describe the evolution of action potentials in the brain, see e.g. [36], or to model interest rates, see [6]. Since analytical solutions to these equations are in general not computable there is a high demand for numerical schemes to approximate these processes.

In this work we are concerned with semilinear parabolic SPDEs on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and on the time interval $[0, T]$ for some $T \in (0, \infty)$ with some filtration $(\mathcal{F}_t)_{t \in [0, T]}$ fulfilling the usual conditions. These SPDEs are of the following general form

$$dX_t = (AX_t + F(X_t)) dt + B(X_t) dW_t, \quad X_0 = \xi. \quad (1)$$

The solution process $(X_t)_{t \in [0, T]}$ is H_γ -valued for some suitable $\gamma \in [0, 1)$ and $(W_t)_{t \in [0, T]}$ is a U -valued Q -Wiener process. Details on the operators, spaces and processes will be given in Section 2.

Even though there was a lot of research on numerical methods for stochastic differential equations in infinite dimensions in the last years, for example [1, 10, 11, 13, 15, 16, 24, 26, 27, 28, 37, 40] methods with high order of convergence and derivative-free schemes remain rare, see [2, 3, 4, 5, 9, 20] or [38]. The numerical approximation of SPDEs requires the discretization of both the time and space domain as well as the Hilbert space-valued stochastic process. When it comes to space most schemes work with a spectral Galerkin method or a finite element discretization to obtain a finite dimensional system of stochastic differential equations, see e.g. [1, 20, 23, 37] or [40]. Concerning the approximation with respect to the temporal direction the linear implicit Euler method is the benchmark, see [8, 12, 14] or [22]. Here, we would like to mention the vitally work on higher order methods like the Milstein scheme in [2, 3, 24, 25].

*e-mail: leonhard@math.uni-luebeck.de

†e-mail: roessler@math.uni-luebeck.de

Recently it has been shown by A. Jentzen and P. E. Kloeden [16] that a higher order of convergence can be obtained when employing schemes which are developed on the basis of the mild solution of (1), that is

$$X_t = e^{At} \xi + \int_0^t e^{A(t-s)} F(X_s) ds + \int_0^t e^{A(t-s)} B(X_s) dW_s \quad \text{P-a.s.} \quad (2)$$

for $t \in [0, T]$. Based on this finding the exponential Euler scheme, see [16], the Milstein scheme for SPDEs in [20] or the numerical scheme in [27] have been build. In the present paper, we focus on the Milstein scheme recently proposed by A. Jentzen and M. Röckner [20] and derive a scheme which is free of derivatives, therefore easier to compute and in general more efficient compared to the original Milstein scheme or the linear implicit Euler scheme.

In order to make our main result more clear, we first consider the Milstein scheme for finite dimensional stochastic differential equations (SDEs). Let $n, k \in \mathbb{N}$ and let $(W_t)_{t \in [0, T]}$ be a k -dimensional Brownian motion with respect to $(\mathcal{F}_t)_{t \in [0, T]}$. Furthermore, assume $a: \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $b = (b_1, \dots, b_k): \mathbb{R}^n \rightarrow \mathbb{R}^{n \times k}$ with $b_j(x) = (b_{1,j}(x), \dots, b_{n,j}(x))^T$, $j \in \{1, \dots, k\}$, $x \in \mathbb{R}^n$, to be Lipschitz continuous functions. Then, the n -dimensional system of SDEs

$$dX_t = a(X_t) dt + \sum_{j=1}^k b_j(X_t) dW_t^j,$$

for $t \in [0, T]$ with initial value $X_0 = \xi \in \mathbb{R}^n$ has a unique solution [21]. Let an equidistant discretization of the time interval $[0, T]$ with step size $h = \frac{T}{M}$ for some $M \in \mathbb{N}$ and $t_m = m h$ for $m \in \{0, \dots, M\}$ be given. Further, let $\Delta W_m^j = W_{t_{m+1}}^j - W_{t_m}^j$. Then, the stochastic double integrals can be expressed as

$$\int_{t_m}^{t_{m+1}} \int_{t_m}^s dW_u^j dW_s^i + \int_{t_m}^{t_{m+1}} \int_{t_m}^s dW_u^i dW_s^j = \Delta W_m^i \Delta W_m^j,$$

for $i, j \in \{1, \dots, k\}$ with $i \neq j$, where the right hand side can be easily simulated. Therefore, we assume the SDE to be commutative, i.e.

$$\sum_{r=1}^n b_{r,j} \frac{\partial b_{l,i}}{\partial x_r} = \sum_{r=1}^n b_{r,i} \frac{\partial b_{l,j}}{\partial x_r}$$

for $l \in \{1, \dots, n\}$ and $i, j \in \{1, \dots, k\}$. Then, for the commutative SDE system the Milstein scheme can be reformulated as $Y_0 = \xi$ and

$$\begin{aligned} Y_{m+1}^M &= Y_m^M + h a(Y_m^M) + \sum_{j=1}^k b_j(Y_m^M) \Delta W_m^j + \frac{1}{2} \sum_{i,j=1}^k \left(\frac{\partial b_{l,i}}{\partial x_r}(Y_m^M) \right)_{1 \leq l, r \leq n} b_j(Y_m^M) (\Delta W_m^i \Delta W_m^j) \\ &\quad - \frac{h}{2} \sum_{j=1}^k \left(\frac{\partial b_{l,j}}{\partial x_r}(Y_m^M) \right)_{1 \leq l, r \leq n} b_j(Y_m^M), \end{aligned}$$

for $m \in \{0, \dots, M-1\}$, which is easy to implement because no double integrals have to be simulated, see [21] for more details. Compared to the Euler-Maruyama method having strong order 1/2, the Milstein scheme attains strong order 1.0 in this case. However, for the Jacobian $\left(\frac{\partial b_{l,i}}{\partial x_r}(Y_m^M) \right)_{1 \leq l, r \leq n}$ one has to evaluate n^2 scalar (nonlinear) functions at Y_m^M for $i = 1, \dots, k$ in each time step. Thus, for an approximation at time T one has to evaluate $\mathcal{O}(M n^2 k)$ scalar nonlinear functions only due to the Jacobian matrix. If n and k are moderately large, e.g. $n = k = 30$, already $30^3 = 27000$ function evaluations are necessary for the Jacobian in each step, which needs much computation time. On the other hand, one step of the Euler-Maruyama scheme is much cheaper because for the diffusion b only $30^2 = 900$ scalar (nonlinear) functions have to be evaluated whereas an evaluation of the Jacobian is

not necessary. In general, the Euler-Maruyama scheme needs one evaluation of the drift a and the diffusion b each step which results in only $\mathcal{O}(Mnk)$ evaluations of scalar (nonlinear) functions for an approximation at time T however also with only low order of convergence. This problem is well known and a special technique for efficient schemes overcoming this trade-off in the SDE setting has been introduced by A. R  kler [32, 33, 34]. Especially, in case of commutative noise strong order 1.0 schemes with only $\mathcal{O}(Mnk)$ evaluations of scalar functions are proposed in [34].

In the infinite dimensional setting one has to be much more careful as the number of function evaluations in the Milstein scheme is 'cubic' with respect to the dimension of the finite dimensional projection subspaces. The dimensions N and K of these subspaces have to increase to get higher approximation accuracy. The Milstein scheme for SPDE (1) proposed by A. Jentzen and M. R  ckner [20] reads as $Y_0^{N,K,M} = P_N \xi$ and

$$\begin{aligned} Y_{m+1}^{N,K,M} = & P_N \left(e^{Ah} \left(Y_m^{N,K,M} + hF(Y_m^{N,K,M}) + B(Y_m^{N,K,M}) \Delta W_m^{K,M} \right. \right. \\ & + \frac{1}{2} B'(Y_m^{N,K,M}) (B(Y_m^{N,K,M}) \Delta W_m^{K,M}, \Delta W_m^{K,M}) \\ & \left. \left. - \frac{h}{2} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j B'(Y_m^{N,K,M}) (B(Y_m^{N,K,M}) \tilde{e}_j, \tilde{e}_j) \right) \right), \end{aligned}$$

for $m = 0, 1, \dots, M-1$. Details on the operators and the notation can be found in Section 3. In their examples, A. Jentzen and M. R  ckner [20] solve the issue of high dimensionality by restricting the operator F to be of the form $(F(v))(x) = f(x, v(x))$ and the operator B to be in a class which is multiplicative in the Q -Wiener process, i.e. $(B(v)u)(x) = b(x, v(x)) \cdot u(x)$ for all $x \in (0, 1)^d$, $u, v \in H = U = L^2((0, 1)^d, \mathbb{R})$, $f, b: (0, 1)^d \times \mathbb{R} \rightarrow \mathbb{R}$ and $d = 1, 2, 3$. Thereby, they can avoid computational costs which are 'cubic' in the dimensions of the problem for each step. Moreover, their scheme is also applicable if this restriction does not hold, however then the computational cost also become 'cubic' in the dimensions of the projection subspaces.

On the other hand in [38] a derivative-free version of the Milstein scheme for SPDEs is derived under certain conditions. However, this scheme is not applicable to general equations of type (1) which do not have to be multiplicative in the Q -Wiener process. Their scheme has the following general form

$$\begin{aligned} Y_{m+1}^{N,K,M} = & P_N \left(e^{Ah} \left(Y_m^{N,K,M} + hF(Y_m^{N,K,M}) + B(Y_m^{N,K,M}) \Delta W_m^{K,M} \right. \right. \\ & \left. \left. + \frac{1}{2} BB(Y_m^{N,K,M}, h) (\Delta W_m^{K,M}, \Delta W_m^{K,M}) - \frac{h}{2} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j BB(Y_m^{N,K,M}, h) (\tilde{e}_j, \tilde{e}_j) \right) \right), \end{aligned}$$

where $m = 0, 1, \dots, M-1$. In [38] the bilinear approximation operator BB is given for examples which are multiplicative in the Q -Wiener process only. The operator $BB(v, h)$ is the approximation of $B'(v)B(v)$ and has to fulfill the following two assumptions such that the rate of convergence of the Milstein scheme is maintained: There exists a constant C , independent of $h > 0$, such that

$$\|BB(v, h) - BB(w, h)\|_{L_{HS}^{(2)}(U_0, H)}^2 \leq \frac{C}{h} \|v - w\|_H^2 \quad (3)$$

$$\|BB(v, h) - B'(v)B(v)\|_{L_{HS}^{(2)}(U_0, H)}^2 \leq Ch \left(1 + \|v\|_{H_\beta}^4 \right) \quad (4)$$

for all $v, w \in H_\beta$ and some $\beta \in [0, 1)$.

Remark 1.1. These assumptions do not have to be and are not fulfilled by the scheme that we propose in the following because we neither need to assume the relation $B(v(x))u(x) = b(x, v(x)) \cdot u(x)$, for all $x \in (0, 1)^d$, $v \in H_\beta$, $u \in U_0$ nor do we assume that the approximations $BB(v, h)(\cdot, \cdot)$ of the operators $B'(v)B(v)(\cdot, \cdot)$ have to be bilinear operators.

Here we demonstrate a different approach to deal with the problem of high dimensionality which is able to treat equations which are not restricted to be multiplicative in the Q -Wiener process and free of derivatives. For the special case of multiplicative noise, our new approach has the same effective order of convergence as the schemes proposed in [20] or [38] since the computational cost is of the same order of magnitude. However, to treat this class is not our main goal and in the general case we can improve the effective order of convergence compared to the Milstein scheme in [20]. Recently, a special technique to reduce the computational costs by a factor depending on the dimensions of the considered SDE system to be solved has been proposed for the first time by Rößler for finite dimensional SDEs, see e.g. [32, 33, 34]. This technique opened the door for the efficient application of higher order schemes in the case of high dimensional SDE systems. Here, the idea is to carry over this approach due to Rößler to the infinite dimensional setting of SPDEs where it becomes even more powerful because one can achieve an improvement of the order of convergence. In this work we will derive a scheme which is efficiently applicable to a broad class of SPDEs. A more flexible choice of the operator approximating the derivative is possible as the assumptions (3)–(4) needed in [38] do not have to be fulfilled. We approximate the derivative and reduce the large number of function evaluations by choosing the approximation operator carefully. The resulting scheme, the efficient derivative-free Milstein scheme, approximates the mild solution (2) of (1) with the same theoretical order of convergence with respect to the spatial and time discretizations as the schemes by [20] and, in case of multiplicative noise, the scheme by [38]. However, the computational cost is reduced by one order of magnitude for a general class of commutative semilinear SPDEs and the overall order of convergence can thus be increased.

2 Framework for the considered SPDEs

$$\mathrm{tr}(Q) = \sum_{j \in \mathcal{J}} \langle Q \tilde{e}_j, \tilde{e}_j \rangle_U < \infty$$

$$\begin{aligned} dX_t &= (AX_t + F(X_t)) dt + B(X_t) dW_t, \quad t \in (0, T] \\ X_0 &= \xi \end{aligned} \tag{5}$$

For the analysis of convergence for the efficient derivative-free Milstein scheme we make the following assumptions which are similar as for the original Milstein scheme proposed in [20]. We can dispense with the assumptions on the second derivative of the drift coefficient F however. For easy comparison of the presented results, we adopt the notation used in [20]:

(A1) For the linear operator $A: D(A) \subset H \rightarrow H$ there exist eigenfunctions $(e_i)_{i \in \mathcal{I}}$ in H and eigenvalues $(\lambda_i)_{i \in \mathcal{I}}$ with $\lambda_i \in (0, \infty)$ and $\inf_{i \in \mathcal{I}} \lambda_i > 0$, such that $-Ae_i = \lambda_i e_i$ for all $i \in \mathcal{I}$, where \mathcal{I} is a finite or countable index set, and such that the eigenfunctions constitute an orthonormal basis of H . The domain of A is defined as $D(A) = \{u \in H : \sum_{i \in \mathcal{I}} |\lambda_i|^2 |\langle u, e_i \rangle_H|^2 < \infty\}$ and for all $x \in D(A)$

$$Ax = \sum_{i \in \mathcal{I}} -\lambda_i \langle x, e_i \rangle_H e_i.$$

Here, A is the generator of a C_0 -semigroup $\{S(t) : t \geq 0\}$ of linear operators in H which are denoted as $S(t) = e^{At}$ for $t \geq 0$ [31]. For $\rho \in [0, \infty)$ let $H_\rho := D((-A)^\rho)$ with norm $\|u\|_{H_\rho} := \|(-A)^\rho u\|_H$ for $u \in H_\rho$ denote the interpolation spaces which are real Hilbert spaces of domains of fractional powers of $-A: D(A) \rightarrow H$ with the relation $H_{\rho_2} \subset H_{\rho_1} \subset H$ for $\rho_2 \geq \rho_1 \geq 0$ [35].

(A2) Let $F: H_\beta \rightarrow H$ for some $\beta \in [0, 1)$ and we assume the mapping to be continuously Fréchet differentiable with $\sup_{v \in H_\beta} \|F'(v)\|_{L(H)} < \infty$.

(A3) Let $B: H_\beta \rightarrow L_{HS}(U_0, H)$, we assume B to be twice continuously Fréchet differentiable with $\sup_{v \in H_\beta} \|B'(v)\|_{L(H, L(U, H))} < \infty$, $\sup_{v \in H_\beta} \|B''(v)\|_{L^{(2)}(H, L(U, H))} < \infty$ and $B(H_\delta) \subset L_{HS}(U_0, H_\delta)$ for some $\delta \in (0, \frac{1}{2})$. Furthermore, there exists a constant $C > 0$ such that

$$\begin{aligned} \|B(u)\|_{L(U, H_\delta)} &\leq C(1 + \|u\|_{H_\delta}), \\ \|B'(v)B(v) - B'(w)B(w)\|_{L_{HS}^{(2)}(U_0, H)} &\leq C\|v - w\|_H, \\ \|(-A)^{-\vartheta} B(v)Q^{-\alpha}\|_{L_{HS}(U_0, H)} &\leq C(1 + \|v\|_{H_\gamma}) \end{aligned}$$

for all $u \in H_\delta$, $v, w \in H_\gamma$ with parameters $\alpha \in (0, \infty)$, $\delta, \vartheta \in (0, \frac{1}{2})$ and $\gamma \in [\max(\beta, \delta), \delta + \frac{1}{2}]$. Thus, it follows that $\beta \in [0, \delta + \frac{1}{2}]$. Here, let $L^{(2)}(H, L(U, H)) = L(H, L(H, L(U, H)))$ and let for all $v \in H_\beta$ the mapping $B'(v)B(v): U_0 \times U_0 \rightarrow H$ with $(B'(v)B(v))(u, \tilde{u}) = (B'(v)(B(v)u))\tilde{u}$ for $u, \tilde{u} \in U_0$ be a bilinear Hilbert-Schmidt operator in $L_{HS}^{(2)}(U_0, H) = L_{HS}(U_0, L_{HS}(U_0, H))$. Moreover, for all $v \in H_\beta$ the operator $B'(v)B(v) \in L_{HS}^{(2)}(U_0, H)$ is assumed to be symmetric, i.e., the operator fulfills the commutativity condition

$$(B'(v)(B(v)u))\tilde{u} = (B'(v)(B(v)\tilde{u}))u \quad (6)$$

for all $u, \tilde{u} \in U_0$.

(A4) The initial value $\xi: \Omega \rightarrow H_\gamma$ is assumed to be a $\mathcal{F}_0\text{-}\mathcal{B}(H_\gamma)$ -measurable random variable such that $E[\|\xi\|_{H_\gamma}^4] < \infty$ is fulfilled.

Note, that $\sup_{v \in H_\beta} \|B'(v)\|_{L(H, L_{HS}(U_0, H))} \leq \text{tr}(Q) \sup_{v \in H_\beta} \|B'(v)\|_{L(H, L(U, H))} < \infty$ and since H_β is a dense subset of H it follows that $B: H_\beta \rightarrow L_{HS}(U_0, H)$ can be continuously extended to a globally Lipschitz continuous mapping $\tilde{B}: H \rightarrow L_{HS}(U_0, H)$. However, to keep the presentation simple it is not distinguished between B and \tilde{B} in the following.

If the assumptions (A1)–(A4) are fulfilled, then for the SPDE (5) there exists a unique mild solution, see A. Jentzen and M. Röckner [19, 20].

Proposition 2.1 (Existence and uniqueness of the mild solution). *Let assumptions (A1)–(A4) be fulfilled. Then, there exists an up to modifications unique predictable mild solution $X : [0, T] \times \Omega \rightarrow H_\gamma$ for (5) with $\sup_{t \in [0, T]} \mathbb{E}[\|X_t\|_{H_\gamma}^4 + \|B(X_t)\|_{LHS(U_0, H_\delta)}] < \infty$ and*

$$X_t = e^{At}\xi + \int_0^t e^{A(t-s)}F(X_s)ds + \int_0^t e^{A(t-s)}B(X_s)dW_s \quad \text{P-a.s.} \quad (7)$$

for all $t \in [0, T]$ with

$$\sup_{\substack{s, t \in [0, T] \\ s \neq t}} \frac{(E[\|X_t - X_s\|_{H_r}^p])^{\frac{1}{p}}}{|t - s|^{\min(\gamma-r, \frac{1}{2})}} < \infty$$

for every $r \in [0, \gamma]$. Further, the process $(X_t)_{t \in [0, T]}$ is continuous with respect to $(\mathbb{E}[\|\cdot\|_{H_\gamma}^4])^{1/4}$.

3 An efficient derivative-free Milstein scheme

To derive a numerical scheme for SPDEs we project the infinite dimensional state space onto a finite dimensional subspace and discretize the time interval. In the following, let $(\mathcal{I}_N)_{N \in \mathbb{N}}$ and $(\mathcal{J}_K)_{K \in \mathbb{N}}$ be sequences of finite subsets such that $\mathcal{I}_N \subset \mathcal{I}$ and $\mathcal{J}_K \subset \mathcal{J}$ for all $K, N \in \mathbb{N}$. For $N \in \mathbb{N}$, let $P_N : H \rightarrow H_N$ denote the projection of the infinite dimensional space H onto the finite dimensional subspace $H_N = \text{span}\{e_i : i \in \mathcal{I}_N\} \subset H$ defined by

$$P_N v = \sum_{i \in \mathcal{I}_N} \langle v, e_i \rangle_H e_i$$

for $v \in H$. Analogously, for $K \in \mathbb{N}$ let $(W_t^K)_{t \in [0, T]}$ denote the projection of the U -valued Q -Wiener process $(W(t))_{t \in [0, T]}$ onto the finite dimensional subspace $U_K = \text{span}\{\tilde{e}_j : j \in \mathcal{J}_K\} \subset U$ defined by

$$W_t^K = \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \langle W_t, \tilde{e}_j \rangle_U \tilde{e}_j = \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \sqrt{\eta_j} \beta_t^j \tilde{e}_j \quad \text{P-a.s.}$$

where $(\beta_t^j)_{t \in [0, T]}$ are independent real-valued Brownian motions for $j \in \mathcal{J}_K$ with $\eta_j \neq 0$. As the next step, we consider a discretization of the time. For legibility, the interval $[0, T]$ is divided into $M \in \mathbb{N}$ equally spaced subsets of length $h = \frac{T}{M}$ with $t_m = m h$ for $m = 0, \dots, M$. Especially, we make use of the increments

$$\Delta W_m^{K, M} := W_{t_{m+1}}^K - W_{t_m}^K = \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \sqrt{\eta_j} \Delta \beta_m^j \tilde{e}_j \quad \text{P-a.s.}$$

with $\Delta \beta_m^j = \beta_{t_{m+1}}^j - \beta_{t_m}^j$ P-a.s. for $m \in \{0, 1, \dots, M-1\}$. We assume commutativity as stated in assumption (A3) which allows to rewrite

$$\begin{aligned} & e^{A(T-t)} \int_t^T B'(X_t) \left(\int_t^s B(X_t) dW_r^K \right) dW_s^K \\ &= e^{A(T-t)} \left(\frac{1}{2} B'(X_t) (B(X_t)(W_T^K - W_t^K), (W_T^K - W_t^K)) - \frac{T-t}{2} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j B'(X_t) (B(X_t) \tilde{e}_j, \tilde{e}_j) \right), \end{aligned}$$

for $t \in [0, T]$ such that the iterated stochastic integral can be split into two parts and simulation becomes straight forward (see [20] for a proof).

For some arbitrarily fixed N, K and M , let $(Y_m^{N, K, M})_{0 \leq m \leq M}$ with \mathcal{F}_{t_m} - $\mathcal{B}(H)$ -measurable random variables $Y_m^{N, K, M} : \Omega \rightarrow H_N$ denote the discrete time approximation process for $(X_{t_m})_{0 \leq m \leq M}$. Now we introduce a scheme which does not employ the derivative of B and therefore allows for the efficient application to a broader class of SPDEs than the Milstein scheme proposed in [20]. Following ideas

for ordinary SDEs in [34] we propose a scheme which is characterized by moving one of the sums into the argument. Thereby, fewer function evaluations are necessary which results in a higher order of convergence. For some arbitrarily fixed N , K and M we define the efficient derivative-free Milstein scheme as $Y_0^{N,K,M} = P_N \xi$ and

$$\begin{aligned} Y_{m+1}^{N,K,M} = & P_N \left(e^{Ah} \left(Y_m^{N,K,M} + hF(Y_m^{N,K,M}) + B(Y_m^{N,K,M}) \Delta W_m^{K,M} \right. \right. \\ & + \frac{1}{\sqrt{h}} \left(B \left(Y_m^{N,K,M} + \frac{1}{2} \sqrt{h} P_N B(Y_m^{N,K,M}) \Delta W_m^{K,M} \right) - B(Y_m^{N,K,M}) \right) \Delta W_m^{K,M} \\ & \left. \left. + \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \bar{B}(Y_m^{N,K,M}, h, j) \right) \right) \end{aligned} \quad (8)$$

for $m = 0, 1, \dots, M-1$ where in general we choose

$$\bar{B}(Y_m^{N,K,M}, h, j) = B \left(Y_m^{N,K,M} - \frac{h}{2} \sqrt{\eta_j} P_N B(Y_m^{N,K,M}) \tilde{e}_j \right) \sqrt{\eta_j} \tilde{e}_j - B(Y_m^{N,K,M}) \sqrt{\eta_j} \tilde{e}_j.$$

It is important to note that the proposed efficient derivative-free Milstein scheme uses a special approximation of the derivative in the original Milstein scheme which turns out to be very efficient. Especially, approximating the derivative in the way it is done in the efficient derivative-free Milstein scheme does not influence the error estimate significantly. Apart from constants it can be proved to be the same as for the Milstein scheme. The main result of this article is given as follows:

Theorem 3.1. *Let assumptions (A1)–(A4) be fulfilled. Then, there exists a constant $C \in (0, \infty)$ independent of N , K and M such that for $(Y_m^{N,K,M})_{0 \leq m \leq M}$ defined by the efficient derivative-free Milstein scheme in (8) it holds*

$$\left(\mathbb{E} \left[\|X_{t_m} - Y_m^{N,K,M}\|_H^2 \right] \right)^{\frac{1}{2}} \leq C \left(\left(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \right)^{-\gamma} + \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^\alpha + M^{-\min(2(\gamma-\beta), \gamma)} \right)$$

for all $m \in \{0, 1, \dots, M\}$ and all $N, K, M \in \mathbb{N}$.

For the proof of Theorem 3.1 we refer to Section 5.

Thus, under very similar assumptions (A1)–(A4) as for the Milstein scheme in [20] it is possible to prove the same order of convergence for the efficient derivative-free Milstein scheme. Moreover, as for the Milstein scheme it should be straight forward to approximate the exponential term e^{At} by $(I - At)^{-1}$, $t \in [0, T]$, see [9].

4 Computational costs and effective order of convergence

Convergence results where the order of convergence depends directly on the sets \mathcal{I}_N , \mathcal{J}_K and on the parameter M like in Theorem 3.1 are important to understand the dependence of the error on the dimensionality of the approximation space. However, in order to judge the quality of an algorithm, we are mainly interested in its error and cost. That is why it is important to consider the order of convergence with respect to the computational costs, i.e. errors versus computational costs, which we call the effective order of convergence, see also [34]. Since measured computation time may depend on the implementation of an algorithm, an established theoretical cost model as in [39] is applied to be more objective.

Let V be a real vector space. If $v \in V$ is part of the considered problem to be solved, then an algorithm needs some information about v which can be seen as an oracle or as a black box. As (linear) information we consider the evaluation of any (linear) functional $\phi: V \rightarrow \mathbb{R}$ and denote the space of such functionals as V^* . Clearly, evaluating $\phi \in V^*$ produces some computational cost, say $\text{cost}(\phi) = c > 0$. Typically, each arithmetic operation or evaluation of sine, cosine, the exponential

function etc. produces cost of one unit whereas the evaluation of a functional ϕ produces cost $c \gg 1$. Assuming $c \gg 1$, the informational cost will dominate the cost for arithmetic operations in the algorithm. That is why we concentrate on the costs for evaluating functionals $\phi \in V^*$, see also e.g. [39]. Typical examples in case of a Hilbert space V are $\phi_i(v) = \langle v, u_i \rangle_V$ for some $u_i \in V$, $i = 1, \dots, n$ with $\text{cost}(\phi_1, \dots, \phi_n) = cn$. Moreover, if V is the space of mappings $f: H \rightarrow \mathbb{R}$ then one can consider the dirac functional $\delta_x \in V^*$ with $\delta_x f = f(x)$ for some $x \in H$. So, for $x_1, \dots, x_n \in H$ one can get the function evaluations of $f(x_1), \dots, f(x_n)$ with $\text{cost}(\delta_{x_1}, \dots, \delta_{x_n}) = cn$. In addition, we assume that each independent realisation of a $N(0, 1)$ -distributed random variable can be simulated with cost one.

Assume that, e.g., $|\mathcal{I}_N| = N$, $|\mathcal{J}_K| = K$ and that $\eta_j \neq 0$ for all $j \in \mathcal{J}_K$ for all $K, N \in \mathbb{N}$ which is the worst case for the computational effort. For an implementation of the considered algorithms it is usual to identify H_N by \mathbb{R}^N applying the natural isomorphism $\pi: H_N \rightarrow \mathbb{R}^N$ with $\pi(v) = (\langle v, e_i \rangle)_{1 \leq i \leq N}$ for $v \in H_N$ and analogously we identify U_K by \mathbb{R}^K . Let $y, v \in H_N$, $u \in U_K$, $L(H, E)_N = \{T|_{H_N} : T \in L(H, E)\}$ for some vector space E and let $L_{HS}(U, H)_{K,N} = \{P_N T|_{U_K} : T \in L_{HS}(U, H)\}$. Then, we obtain the following computational cost:

- i) One evaluation of the mapping $P_N \circ F: H \rightarrow H_N$ with $P_N F(y) = \sum_{i \in \mathcal{I}_N} \langle F(y), e_i \rangle_H e_i$ is determined by the functionals $\langle F(y), e_i \rangle_H$ for $i \in \mathcal{I}_N$ with $\text{cost}(P_N F(y)) = cN$.
- ii) Due to $P_N B(y)u = \sum_{i \in \mathcal{I}_N} \sum_{j \in \mathcal{J}_K} \langle B(y)\tilde{e}_j, e_i \rangle_H \langle u, \tilde{e}_j \rangle_U e_i$ one evaluation of $P_N \circ B(\cdot)|_{U_K}: H \rightarrow L_{HS}(U, H)_{K,N}$ needs the evaluation of the functionals $\langle B(y)\tilde{e}_j, e_i \rangle_H$ for $i \in \mathcal{I}_N$ and $j \in \mathcal{J}_K$ with $\text{cost}(P_N \circ B(y)|_{U_K}) = cNK$.
- iii) Finally, observe that $P_N((B'(y)v)u) = \sum_{k,l \in \mathcal{I}_N} \sum_{j \in \mathcal{J}_K} \langle (B'(y)e_k)\tilde{e}_j, e_l \rangle_H \langle v, e_k \rangle_H \langle u, \tilde{e}_j \rangle_U e_l$ holds for $P_N \circ B'(\cdot)(\cdot, \cdot)|_{H_N, U_K}: H \rightarrow L(H, L_{HS}(U, H)_{K,N})_N$ with $\text{cost}(P_N \circ B'(y)(\cdot, \cdot)|_{H_N, U_K}) = cN^2K$ since the functionals $\langle (B'(y)e_k)\tilde{e}_j, e_l \rangle_H$ have to be evaluated for all $k, l \in \mathcal{I}_N$ and $j \in \mathcal{J}_K$.

Provided that for $T \in L_{HS}(U, H)_{K,N}$ all functionals $\langle T\tilde{e}_j, e_i \rangle_H$ and $\langle u, \tilde{e}_j \rangle_U$ are known for $i \in \mathcal{I}_N$ and $j \in \mathcal{J}_K$, then $Tu = \sum_{i \in \mathcal{I}_N} \sum_{j \in \mathcal{J}_K} \langle u, \tilde{e}_j \rangle_U \langle T\tilde{e}_j, e_i \rangle_H e_i$ and the calculation of $\pi(Tu)_i = \langle Tu, e_i \rangle_H$ needs K multiplications and $K - 1$ summations for each $i \in \mathcal{I}_N$ and thus $\text{cost}(\pi(Tu)) = 2NK - 1$. Analogously, for $T \in L(H, L_{HS}(U, H)_{K,N})_N$ it follows that $\text{cost}(\pi((Tv)u)) = 3N^2K - 1$ provided that the functionals $\langle (Te_k)\tilde{e}_j, e_l \rangle_H$, $\langle v, e_k \rangle_H$ and $\langle u, \tilde{e}_j \rangle_U$ are known for all $k, l \in \mathcal{I}_N$ and $j \in \mathcal{J}_K$.

To assess the usefulness and efficiency of the proposed efficient derivative-free Milstein scheme denoted as (DFM) we will compare it to the Milstein scheme [20] denoted as (MIL), the linear implicit Euler scheme considered in, e.g., [22, 37] denoted as (LIE) and the exponential Euler scheme, see e.g. [17, 27], denoted as (EES). Here, we want to mention that the Runge-Kutta type scheme proposed in [38] is not taken into account because it can not be applied to the general class of SPDEs under consideration.

The computational cost of the Milstein scheme for each time step are determined by one evaluation of $P_N \circ F$, $P_N \circ B(\cdot)|_{U_K}$ and one evaluation of $P_N \circ B'(\cdot)|_{H_N, U_K}$. In addition the following linear and bilinear operators have to be applied: One application of $P_N \circ B(Y_m^{N,K,M})|_{U_K} \in L_{HS}(U, H)_{K,N}$ (here, calculating $P_N B(Y_m^{N,K,M})\tilde{e}_j$ for a basis element $\tilde{e}_j \in U_K$ is for free because it is the j -th column of the matrix representation $P_N B(Y_m^{N,K,M})|_{U_K} = (b_{i,j}(Y_m^{N,K,M}))_{i \in \mathcal{I}_N, j \in \mathcal{J}_K}$ with $b_{i,j}(Y_m^{N,K,M}) = \langle B(Y_m^{N,K,M})\tilde{e}_j, e_i \rangle_H$ which is already determined), one application of the bilinear operator $P_N \circ B'(Y_m^{N,K,M})|_{H_N, U_K} \in L(H, L_{HS}(U, H)_{K,N})_N$, one application of a operator of type $P_N \circ B'(Y_m^{N,K,M})(v)|_{U_K} \in L_{HS}(U, H)_{K,N}$ (here again the application of the operator to a basis element $\tilde{e}_j \in U_K$ is for free) and one application of $P_N \circ e^{Ah}|_{H_N}: H_N \rightarrow H_N$. In addition, K independent realizations of $N(0, 1)$ -distributed random variables have to be simulated. Summing up, the computational cost for the approximation of one realization of the solution X_T by the Milstein scheme are $\text{cost}(\text{MIL}(N, K, M)) = \mathcal{O}(N^2KM)$.

The introduced efficient derivative-free Milstein scheme (DFM) needs for each time step the evalu-

Scheme	computational costs for evaluation of			# of $N(0, 1)$ r. v.
	$P_N F(\cdot) _{H_N}$	$P_N B(\cdot) _{U_K}$	$P_N B'(\cdot) _{H_N, U_K}$	
Milstein	N	KN	KN^2	K
Lin Euler	N	KN	—	K
Exp Euler	N	KN	—	K
DFM	N	$3KN$	—	K

Table 1: Number of real-valued nonlinear function evaluations and independent $N(0, 1)$ -distributed random variables for each time step.

ation of $P_N \circ F$, two times $P_N \circ B(\cdot)|_{U_K}$ and the evaluation of

$$\sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} P_N B \left(Y_m^{N, K, M} - \frac{h}{2} \sqrt{\eta_j} P_N B(Y_m^{N, K, M}) \tilde{e}_j \right) \sqrt{\eta_j} \tilde{e}_j. \quad (9)$$

Observe that for each $j \in \mathcal{J}_K$ the computation of $P_N B(Y_m^{N, K, M} - \frac{h}{2} \sqrt{\eta_j} P_N B(Y_m^{N, K, M}) \tilde{e}_j) \sqrt{\eta_j} \tilde{e}_j$ results in the calculation of the functionals $\phi_i^j = \langle B(Y_m^{N, K, M} - \frac{h}{2} \sqrt{\eta_j} P_N B(Y_m^{N, K, M}) \tilde{e}_j) \sqrt{\eta_j} \tilde{e}_j, e_i \rangle_H$ for $i \in \mathcal{I}_N$ with $\text{cost}(\phi_1^j, \dots, \phi_N^j) = cN$. Therefore, the evaluation of (9) can be done with cost cNK . In addition, the linear operators $P_N \circ e^{Ah}|_{H_N}: H_N \rightarrow H_N$, $P_N \circ B(Y_m^{N, K, M})|_{U_K} \in L(U, H)_{K, N}$ (note again that calculating $P_N B(Y_m^{N, K, M}) \tilde{e}_j$ for a basis element $\tilde{e}_j \in U_K$ is for free) and $P_N \circ B(Y_m^{N, K, M} + \frac{1}{2} \sqrt{h} P_N B(Y_m^{N, K, M}) \Delta W_m^{K, M})|_{U_K} \in L(U, H)_{K, N}$ have to be applied. Finally, K independent realizations of $N(0, 1)$ -distributed random variables have to be simulated in each step. Thus, the total computational cost for M time steps of the efficient derivative-free Milstein scheme for the approximation of one realization of X_T are $\text{cost}(\text{DFM}(N, K, M)) = \mathcal{O}(NKM)$.

Although both schemes (MIL) and (DFM) have the same order of convergence with respect to the dimensions N , K and M of the finite-dimensional subspaces, their computational costs depend on these parameters with different powers. In contrast to the setting of finite dimensional SDEs with fixed dimensions, for SPDEs on infinite dimensional spaces the dimensions of the finite dimensional projection subspaces have to increase as the accuracy of the approximation increases. Thus, the computational costs do not only depend on M but also on the variable dimensions N and K . Especially, the reduction of the power of N in the computational costs results in an improvement of the order of convergence if one considers errors versus computational costs. Here, we have to point out that computational costs of order $\mathcal{O}(NKM)$ are in some sense optimal within the class of one-step approximations methods because in general one evaluation of the nonlinear diffusion $P_N \circ B(\cdot)|_{U_K}$ already produces computational cost of order $\mathcal{O}(NK)$ for each time step. Further, the linear implicit Euler scheme (LIE) as well as the exponential Euler scheme (EES) have computational costs $\text{cost}(\text{LIE}(N, K, M)) = \text{cost}(\text{EES}(N, K, M)) = \mathcal{O}(NKM)$ which are of the same order as for the introduced efficient derivative-free Milstein scheme (DFM).

Next, the effective order of convergence is determined for the schemes under consideration. First, one has to solve an optimization problem for the optimal choice of the parameters N , K and M such that the error is minimized under the constraint that the computational cost are arbitrarily fixed. Here, one needs to know about the relationship between $\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i$ and $\dim(H_N)$ as well as between $\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j$ and $\dim(U_K)$ for any $N, K \in \mathbb{N}$. Therefore, as an example, we assume that $\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i = \mathcal{O}(N^{\rho_A})$ and $\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j = \mathcal{O}(K^{-\rho_Q})$ for some $\rho_A, \rho_Q > 0$. Moreover, similar results can be obtained under different assumptions as well. Then, for some $q \geq 0$ we investigate the error

$$\text{err}(\text{SCHEME}(N, K, M)) := \left(\mathbb{E} \left[\|X_T - Y_M^{N, K, M}\|_H^2 \right] \right)^{\frac{1}{2}} = \mathcal{O}(N^{-\gamma \rho_A} + K^{-\alpha \rho_Q} + M^{-q}) \quad (10)$$

and minimize $\text{err}(\text{SCHEME}(N, K, M))$ under the constraint that $\text{cost}(\text{SCHEME}(N, K, M)) = \bar{c}$ for some arbitrary constant $\bar{c} > 0$.

For the Milstein scheme (MIL) with $q = \min(2(\gamma - \beta), \gamma)$ and $\text{cost}(\text{MIL}(N, K, M)) = \mathcal{O}(N^2 KM)$ we obtain as an optimal choice

$$N = \mathcal{O}\left(\bar{c}^{\frac{\alpha \rho_Q q}{(2\alpha \rho_Q + \gamma \rho_A)q + \alpha \gamma \rho_A \rho_Q}}\right), \quad K = \mathcal{O}\left(\bar{c}^{\frac{\gamma \rho_A q}{(2\alpha \rho_Q + \gamma \rho_A)q + \alpha \gamma \rho_A \rho_Q}}\right), \quad M = \mathcal{O}\left(\bar{c}^{\frac{\alpha \gamma \rho_A \rho_Q}{(2\alpha \rho_Q + \gamma \rho_A)q + \alpha \gamma \rho_A \rho_Q}}\right),$$

which balances the three summands on the right hand side of (10). As a result of this the effective order of convergence for the error versus the computational costs of the Milstein scheme is

$$\text{err}(\text{MIL}(N, K, M)) = \mathcal{O}\left(\bar{c}^{-\frac{\alpha \gamma \rho_A \rho_Q \min(2(\gamma - \beta), \gamma)}{(2\alpha \rho_Q + \gamma \rho_A) \min(2(\gamma - \beta), \gamma) + \alpha \gamma \rho_A \rho_Q}}\right) \quad (11)$$

which is optimal for the Milstein scheme.

Solving the corresponding optimization problem for the efficient derivative-free Milstein scheme (DFM) with $q = \min(2(\gamma - \beta), \gamma)$ and reduced computational cost given as $\text{cost}(\text{DFM}(N, K, M)) = \mathcal{O}(NKM)$ results in the optimal choice

$$N = \mathcal{O}\left(\bar{c}^{\frac{\alpha \rho_Q q}{(\alpha \rho_Q + \gamma \rho_A)q + \alpha \gamma \rho_A \rho_Q}}\right), \quad K = \mathcal{O}\left(\bar{c}^{\frac{\gamma \rho_A q}{(\alpha \rho_Q + \gamma \rho_A)q + \alpha \gamma \rho_A \rho_Q}}\right), \quad M = \mathcal{O}\left(\bar{c}^{\frac{\alpha \gamma \rho_A \rho_Q}{(\alpha \rho_Q + \gamma \rho_A)q + \alpha \gamma \rho_A \rho_Q}}\right).$$

Then, we get for the efficient derivative-free Milstein scheme (DFM) the effective order of convergence

$$\text{err}(\text{DFM}(N, K, M)) = \mathcal{O}\left(\bar{c}^{-\frac{\alpha \gamma \rho_A \rho_Q \min(2(\gamma - \beta), \gamma)}{(\alpha \rho_Q + \gamma \rho_A) \min(2(\gamma - \beta), \gamma) + \alpha \gamma \rho_A \rho_Q}}\right) \quad (12)$$

which is optimal for the efficient derivative-free Milstein scheme.

It is obvious that the order of the efficient derivative-free Milstein scheme (DFM) is higher than the order of the Milstein scheme (MIL) given in (11). That means, for some arbitrarily prescribed amount of computational costs (or computing time) \bar{c} the minimal possible error $\text{err}(\text{DFM}(N, K, M))$ of the efficient derivative-free Milstein scheme (DFM) decreases with some higher order than the minimal possible error $\text{err}(\text{MIL}(N, K, M))$ of the Milstein scheme as $\bar{c} \rightarrow \infty$.

For the linear implicit Euler scheme (LIE) and the exponential Euler scheme (EES) we obtain the same optimal expressions for N , K and M as for the efficient derivative-free Milstein scheme (DFM). The effective orders of convergence of these schemes are

$$\text{err}(\text{LIE}(N, K, M)) = \mathcal{O}\left(\bar{c}^{-\frac{\alpha \gamma \rho_A \rho_Q q}{(\alpha \rho_Q + \gamma \rho_A)q + \alpha \gamma \rho_A \rho_Q}}\right) \quad (13)$$

and

$$\text{err}(\text{EES}(N, K, M)) = \mathcal{O}\left(\bar{c}^{-\frac{\alpha \gamma \rho_A \rho_Q q}{(\alpha \rho_Q + \gamma \rho_A)q + \alpha \gamma \rho_A \rho_Q}}\right). \quad (14)$$

The parameter $q \geq 0$, however, is in general smaller than for the efficient derivative-free Milstein scheme (DFM) which results in a lower effective order of convergence for the linear implicit Euler scheme (LIE) as well as the exponential Euler scheme (EES).

For the special case of $H = U = L^2((0, 1)^d, \mathbb{R})$ and e.g. Nemytskij operators where $F: H_\beta \rightarrow H$ is given by $(F(v))(x) = f(x, v(x))$ and $B: H_\beta \rightarrow L_{HS}(U_0, H)$ is given by $(B(v)u)(x) = b(x, v(x)) \cdot u(x)$ for some functions $f, b: (0, 1)^d \times \mathbb{R} \rightarrow \mathbb{R}$, $x \in (0, 1)^d$, $v \in H_\beta$, $\beta \in [0, 1]$, $u \in U_0$ and some $d \in \mathbb{N}$, which is the setting also treated in [20] and exclusively in [38], the Milstein scheme can be simplified such that the number of evaluations of the derivative is significantly reduced. In this special setting, the

introduced efficient derivative-free Milstein scheme (8) can be rewritten as $Y_0^{N,K,M} = P_N \xi$ and

$$\begin{aligned} Y_{m+1}^{N,K,M} = & P_N \left(e^{Ah} \left(Y_m^{N,K,M} + hf(\cdot, Y_m^{N,K,M}) + b(\cdot, Y_m^{N,K,M}) \cdot \Delta W_m^{K,M} \right. \right. \\ & + \frac{1}{\sqrt{h}} \left(b(\cdot, Y_m^{N,K,M} + \frac{1}{2}\sqrt{h} P_N b(\cdot, Y_m^{N,K,M}) \cdot \Delta W_m^{K,M}) - b(\cdot, Y_m^{N,K,M}) \right) \cdot \Delta W_m^{K,M} \\ & \left. \left. + \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \bar{B}(Y_m^{N,K,M}, h, j) \right) \right) \end{aligned} \quad (15)$$

for $m = 0, 1, \dots, M-1$. In this case we choose

$$\bar{B}(Y_m, h, j) = \left(b(\cdot, Y_m^{N,K,M} - \frac{h}{2} P_N b(\cdot, Y_m^{N,K,M})) - b(\cdot, Y_m^{N,K,M}) \right) \eta_j \tilde{e}_j$$

for the multiplicative scheme (15), which will be denoted as (DFMM).

For the implementation of scheme (15) one has to compute expressions of the form

$$P_N(f(\cdot, Y_m^{N,K,M}(\cdot))) = \sum_{i \in \mathcal{I}_N} \langle f(\cdot, Y_m^{N,K,M}(\cdot)), e_i \rangle_H e_i = \sum_{i \in \mathcal{I}_N} \left(\int_{(0,1)^d} f(x, Y_m^{N,K,M}(x)) e_i(x) dx \right) e_i$$

where each integral can be approximated by e.g. a standard quadrature formula based on a spatial discretization of $(0,1)^d$. However, the spatial discretization is not in our focus as we restrict our considerations to the time discretization with a general projector P_N independent of the spatial discretization. Then, the computational cost are determined by the calculation of the functionals $\langle f(\cdot, Y_m^{N,K,M}(\cdot)), e_i \rangle_H$ for $i \in \mathcal{I}_N$. Thus $\text{cost}(P_N(f(\cdot, Y_m^{N,K,M}(\cdot)))) = N$. The same applies to the calculation of $P_N(b(\cdot, Y_m^{N,K,M}(\cdot)))$, $P_N(b(\cdot, Y_m^{N,K,M} + \frac{1}{2}\sqrt{h} P_N b(\cdot, Y_m^{N,K,M})))$ and $P_N(b(\cdot, Y_m^{N,K,M} - \frac{h}{2} P_N b(\cdot, Y_m^{N,K,M})))$. Further, the scheme (DFMM) makes use of K independent $N(0,1)$ -distributed random variables. To sum up, the computational costs for the calculation of one approximation of a realisation of X_T for the efficient derivative-free Milstein scheme in this special setting are $\text{cost}(\text{DFMM}(N, K, M)) = \mathcal{O}(MN + MK)$.

The effective order of convergence for the efficient derivative-free Milstein scheme (DFMM) in this special setting can be determined by minimizing the error $\text{err}(\text{DFMM}(N, K, M))$ under the constraint that the costs $\text{cost}(\text{DFMM}(N, K, M)) = \bar{c}$ are arbitrarily fixed. Let $q = \min(2(\gamma - \beta), \gamma)$. Then, a optimal choice is given by

$$N = \mathcal{O}\left(\bar{c}^{\frac{\max(\gamma \rho_A, \alpha \rho_Q)q}{\gamma \rho_A (\max(\gamma \rho_A, \alpha \rho_Q) + q)}}\right), \quad K = \mathcal{O}\left(\bar{c}^{\frac{\max(\gamma \rho_A, \alpha \rho_Q)q}{\alpha \rho_Q (\max(\gamma \rho_A, \alpha \rho_Q) + q)}}\right), \quad M = \mathcal{O}\left(\bar{c}^{\frac{\max(\gamma \rho_A, \alpha \rho_Q)}{\max(\gamma \rho_A, \alpha \rho_Q) + q}}\right).$$

and the effective order of convergence for the error versus the computational costs is

$$\text{err}(\text{DFMM}(N, K, M)) = \mathcal{O}\left(\bar{c}^{-\frac{\max(\gamma \rho_A, \alpha \rho_Q)q}{\max(\gamma \rho_A, \alpha \rho_Q) + q}}\right) \quad (16)$$

which is optimal for the efficient derivative-free Milstein scheme (15). This is the same order as for the Milstein scheme proposed in [20] and for the Runge-Kutta type scheme proposed in [38]. However, like for the Runge-Kutta type scheme in [38], the advantage compared to the Milstein scheme is that no derivative of b has to be calculated. In the following, we do not restrict our analysis to this special case of e.g. Nemytskij operators but allow for a broader class of SPDEs.

5 Proof

The proof of Theorem 3.1 builds on the proof of convergence in [20] - with an addition which accounts for the approximation of the derivative. We estimate

$$\sup_{m \in \{0, \dots, M\}} \left(\mathbb{E} \left[\|X_{t_m} - Y_m^{N,K,M}\|_H^2 \right] \right)^{\frac{1}{2}}.$$

We do not incorporate the analysis of the error which possibly results from the approximation of the coefficients in the spectral projection $P_N X_t = \sum_{n \in \mathcal{I}_N} \langle X_t, e_n \rangle_H e_n$ here. In the proof we use some generic constant C which may change from line to line.

The following elementary facts on the analytical semigroup e^{At} , $t \geq 0$, are employed frequently throughout the proof.

Lemma 5.1 ([29, Lemma 6.13]). *Under assumption (A1) it holds for $t > 0$ and $\theta \in [0, 1]$ that $\|(-A)^\theta e^{At}\|_{L(H)} \leq C_\theta t^{-\theta}$ and $\|(-A)^{-\theta}(e^{At} - I)\|_{L(H)} \leq C_\theta t^\theta$.*

Proof of Theorem 3.1. We use the representation

$$X_{t_m} = e^{At_m} X_0 + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} F(X_s) ds + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B(X_s) dW_s$$

set $Y_m := Y_m^{N,M,K}$ and $W_m^K := W_m^{K,M}$ with $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$, $N \in \mathbb{N}$, $K \in \mathbb{N}$, for legibility and define some auxiliary processes for $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$,

$$\begin{aligned} \bar{X}_{t_m} &:= P_N \left(e^{At_m} X_0 + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} F(X_{t_l}) ds + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} B(X_{t_l}) dW_s^K \right. \\ &\quad \left. + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) dW_s^K \right) \\ \bar{Y}_m &:= P_N \left(e^{At_m} X_0 + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} F(Y_l) ds + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} B(Y_l) dW_s^K \right. \\ &\quad \left. + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} B'(Y_l) \left(\int_{t_l}^s B(Y_l) dW_r^K \right) dW_s^K \right) \\ &= P_N \left(e^{At_m} X_0 + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} F(Y_l) ds + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} B(Y_l) dW_s^K \right. \\ &\quad \left. + \sum_{l=0}^{m-1} e^{A(t_m-t_l)} \left(\frac{1}{2} B'(Y_l) (B(Y_l) \Delta W_l, \Delta W_l) - \frac{h}{2} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j B'(Y_l) (B(Y_l) \tilde{e}_j, \tilde{e}_j) \right) \right). \end{aligned}$$

We estimate

$$\mathbb{E} [\|X_{t_m} - Y_m\|_H^2] = \mathbb{E} [\|X_{t_m} - P_N X_{t_m} + P_N X_{t_m} - \bar{X}_{t_m} + \bar{X}_{t_m} - \bar{Y}_{t_m} + \bar{Y}_{t_m} - Y_m\|_H^2]$$

in several parts

$$\begin{aligned} \mathbb{E} [\|X_{t_m} - Y_m\|_H^2] &\leq 4 \left(\mathbb{E} [\|X_{t_m} - P_N X_{t_m}\|_H^2] + \mathbb{E} [\|P_N X_{t_m} - \bar{X}_{t_m}\|_H^2] \right. \\ &\quad \left. + \mathbb{E} [\|\bar{X}_{t_m} - \bar{Y}_{t_m}\|_H^2] + \mathbb{E} [\|\bar{Y}_{t_m} - Y_m\|_H^2] \right). \end{aligned} \tag{17}$$

The first part is the error that results from the projection of H to a finite dimensional subspace H_N , $N \in \mathbb{N}$. The second and third term arise due to the approximation of the solution process with the Milstein scheme and the last one is the error we obtain by approximating the derivative. After estimating these terms we obtain

$$\mathbb{E} [\|X_{t_m} - Y_m\|_H^2] \leq C_T \left(\left(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \right)^{-2\gamma} + \left(\sup_{j \in \mathcal{I} \setminus \mathcal{I}_K} \eta_j \right)^{2\alpha} + M^{-2 \min(2(\gamma-\beta), \gamma)} \right)$$

$$\begin{aligned}
& + \frac{C_T}{M} \sum_{l=0}^{m-1} \mathbb{E} \left[\|X_{t_l} - Y_l\|_H^2 \right] + C_{T,Q} M^{-2} (\text{tr} Q)^4 \\
& \leq C_{T,Q} \left(\left(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \right)^{-2\gamma} + \left(\sup_{j \in \mathcal{I} \setminus \mathcal{I}_K} \eta_j \right)^{2\alpha} + M^{-2 \min(2(\gamma-\beta), \gamma)} \right)
\end{aligned}$$

by a discrete version of Gronwall's Lemma.

5.1 Spectral Galerkin projection

The error resulting from the spectral Galerkin projection is estimated for $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$, as

$$\begin{aligned}
\mathbb{E} \left[\|X_{t_m} - P_N X_{t_m}\|_H^2 \right] &= \mathbb{E} \left[\|(I - P_N)X_{t_m}\|_H^2 \right] = E \left[\left\| \sum_{n \in \mathcal{I} \setminus \mathcal{I}_N} \langle e_n, X_{t_m} \rangle_H e_n \right\|_H^2 \right] \\
&= E \left[\sum_{n_1, n_2 \in \mathcal{I} \setminus \mathcal{I}_N} \langle e_{n_1}, X_{t_m} \rangle_H \langle e_{n_2}, X_{t_m} \rangle_H \langle e_{n_1}, e_{n_2} \rangle_H \right] \\
&= E \left[\sum_{n \in \mathcal{I} \setminus \mathcal{I}_N} \langle e_n, X_{t_m} \rangle_H^2 \right] = E \left[\sum_{n \in \mathcal{I} \setminus \mathcal{I}_N} \langle e_n, (-A)^{-\gamma} (-A)^\gamma X_{t_m} \rangle_H^2 \right] \\
&= E \left[\sum_{n \in \mathcal{I} \setminus \mathcal{I}_N} \langle (-A)^{-\gamma} e_n, (-A)^\gamma X_{t_m} \rangle_H^2 \right].
\end{aligned}$$

By the series representation of A we obtain further

$$\begin{aligned}
\mathbb{E} \left[\|X_{t_m} - P_N X_{t_m}\|_H^2 \right] &= E \left[\sum_{n \in \mathcal{I} \setminus \mathcal{I}_N} \sum_{k \in \mathcal{I}_N} \lambda_k^{-2\gamma} \langle e_n, e_k \rangle_H^2 \langle e_k, (-A)^\gamma X_{t_m} \rangle_H^2 \right] \\
&= E \left[\sum_{n \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_n^{-2\gamma} \langle e_n, (-A)^\gamma X_{t_m} \rangle_H^2 \right] \\
&\leq E \left[\left(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \right)^{-2\gamma} \sum_{n \in \mathcal{I} \setminus \mathcal{I}_N} \langle e_n, (-A)^\gamma X_{t_m} \rangle_H^2 \right] \\
&\leq \left(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \right)^{-2\gamma} \mathbb{E} \left[\|(-A)^\gamma X_{t_m}\|_H^2 \right] \leq C \left(\inf_{i \in \mathcal{I} \setminus \mathcal{I}_N} \lambda_i \right)^{-2\gamma}
\end{aligned}$$

for $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$, due to (A1)-(A4).

The estimates of the first three terms are not specific to our scheme and the ideas originate from [20]. But we can drop any assumptions on the second derivative of F and generalize the proof a little. The main idea, however, stays the same. For completeness we state the whole proof.

In the following we use

$$\begin{aligned}
\|P_N x\|_H^2 &= \left\| \sum_{n \in \mathcal{I}_N} \langle x, e_n \rangle_H e_n \right\|_H^2 = \sum_{n \in \mathcal{I}_N} \langle x, e_n \rangle_H^2 \\
&\leq \sum_{n \in \mathcal{I}} |\langle x, e_n \rangle_H|^2 = \|x\|_H^2
\end{aligned}$$

several times.

For $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$, we prove

$$\left(\mathbb{E} \left[\|P_N X_{t_m} - \bar{X}_{t_m}\|_H^2 \right] \right)^{\frac{1}{2}} \leq E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} F(X_s) - e^{A(t_m-t_l)} F(X_{t_l}) \right) ds \right\|_H^2 \right]^{\frac{1}{2}}$$

$$\begin{aligned}
& + E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} B(X_s) - e^{A(t_m-t_l)} B(X_{t_l}) \right) dW_s^K \right. \right. \\
& \quad \left. \left. - \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) dW_s^K \right\|_H^2 \right]^{\frac{1}{2}} \\
& + E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B(X_s) (dW_s - dW_s^K) \right\|_H^2 \right]^{\frac{1}{2}} \\
& \leq C_T \left(M^{-\min(2(\gamma-\beta), \gamma)} + \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^\alpha \right),
\end{aligned}$$

$$E \left[\|\bar{X}_{t_m} - \bar{Y}_m\|_H^2 \right] \leq \frac{C_T}{M} \sum_{l=0}^{m-1} E \left[\|X_{t_l} - Y_l\|_H^2 \right] \quad (18)$$

and

$$E \left[\|\bar{Y}_m - Y_m\|_H^2 \right] \leq C_T h^2 (tr Q)^4$$

separately.

5.2 Temporal discretization - nonlinearity F

Next we prove the error resulting from the temporal discretization of the Bochner integral by partitioning the error into three components which we again estimate separately. Let $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$. We show

$$\begin{aligned}
& \left(E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} F(X_s) - e^{A(t_m-t_l)} F(X_{t_l}) \right) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& \leq \left(E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (F(X_s) - F(X_{t_l})) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& \quad + \left(E \left[\left\| \sum_{l=0}^{m-2} \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} - e^{A(t_m-t_l)} \right) F(X_{t_l}) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& \quad + \left(E \left[\left\| \int_{t_{m-1}}^{t_m} \left(e^{A(t_m-s)} - e^{A(t_m-t_{m-1})} \right) F(X_{t_{m-1}}) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& \leq C_T M^{-\gamma}.
\end{aligned}$$

For the first term we obtain by the triangle inequality and the representation of the mild solution $(X_t)_{t \in [0, T]}$

$$\begin{aligned}
& \left(E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (F(X_s) - F(X_{t_l})) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& = \left(E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} \left(\int_0^1 F'(X_{t_l} + r(X_s - X_{t_l}))(X_s - X_{t_l}) dr \right) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& \leq \sum_{l=0}^{m-1} E \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} \left(\int_0^1 F'(X_{t_l} + r(X_s - X_{t_l})) (e^{A(s-t_l)} - I) X_{t_l} dr \right) ds \right\|_H^2 \right]^{\frac{1}{2}}
\end{aligned}$$

$$\begin{aligned}
& + \sum_{l=0}^{m-1} \mathbb{E} \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} \left(\int_0^1 F'(X_{t_l} + r(X_s - X_{t_l})) \left(\int_{t_l}^s e^{A(s-u)} F(X_u) du \right) dr \right) ds \right\|_H^2 \right]^{\frac{1}{2}} \\
& + \left(\sum_{l=0}^{m-1} \mathbb{E} \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} \left(\int_0^1 F'(X_{t_l} + r(X_s - X_{t_l})) \left(\int_{t_l}^s e^{A(s-u)} B(X_u) dW_u \right) dr \right) ds \right\|_H^2 \right] \right)^{\frac{1}{2}}.
\end{aligned}$$

Employing Hölders inequality yields

$$\begin{aligned}
& \left(E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (F(X_s) - F(X_{t_l})) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& \leq \sum_{l=0}^{m-1} E \left[h \int_{t_l}^{t_{l+1}} \left\| e^{A(t_m-s)} \left(\int_0^1 F'(X_{t_l} + r(X_s - X_{t_l})) (e^{A(s-t_l)} - I) X_{t_l} dr \right) \right\|_H^2 ds \right]^{\frac{1}{2}} \\
& + \sum_{l=0}^{m-1} E \left[h \int_{t_l}^{t_{l+1}} \left\| e^{A(t_m-s)} \left(\int_0^1 F'(X_{t_l} + r(X_s - X_{t_l})) \left(\int_{t_l}^s e^{A(s-u)} F(X_u) du \right) dr \right) \right\|_H^2 ds \right]^{\frac{1}{2}} \\
& + \left(\sum_{l=0}^{m-1} \mathbb{E} \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} \left(\int_0^1 F'(X_{t_l} + r(X_s - X_{t_l})) \left(\int_{t_l}^s e^{A(s-u)} B(X_u) dW_u \right) dr \right) ds \right\|_H^2 \right] \right)^{\frac{1}{2}}
\end{aligned}$$

and further by the Itô isometry, (A1) and Lemma 5.1

$$\begin{aligned}
& \left(E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (F(X_s) - F(X_{t_l})) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& \leq C \sum_{l=0}^{m-1} E \left[h \int_{t_l}^{t_{l+1}} \int_0^1 \|F'(X_{t_l} + r(X_s - X_{t_l}))\|_{L(H)}^2 \underbrace{\|(-A)^{-\gamma} (e^{A(s-t_l)} - I)\|_{L(H)}^2}_{\leq (s-t_l)^{2\gamma}} \|X_{t_l}\|_{H_\gamma}^2 dr ds \right]^{\frac{1}{2}} \\
& + C \sum_{l=0}^{m-1} E \left[h \int_{t_l}^{t_{l+1}} \int_0^1 \|F'(X_{t_l} + r(X_s - X_{t_l}))\|_{L(H)}^2 \left\| \left(\int_{t_l}^s e^{A(s-u)} F(X_u) du \right) \right\|_H^2 dr ds \right]^{\frac{1}{2}} \\
& + \left(C \sum_{l=0}^{m-1} h E \left[\int_{t_l}^{t_{l+1}} \int_0^1 \|F'(X_{t_l} + r(X_s - X_{t_l}))\|_{L(H)}^2 \left\| \left(\int_{t_l}^s e^{A(s-u)} B(X_u) dW_u \right) \right\|_H^2 dr ds \right] \right)^{\frac{1}{2}}.
\end{aligned}$$

By (A1)-(A4) we get

$$\begin{aligned}
& \left(E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (F(X_s) - F(X_{t_l})) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
& \leq CMh^{1+\gamma} + C \sum_{l=0}^{m-1} E \left[h \int_{t_l}^{t_{l+1}} \left\| \left(\int_{t_l}^s e^{A(s-u)} F(X_u) du \right) \right\|_H^2 ds \right]^{\frac{1}{2}} \\
& + \left(C \sum_{l=0}^{m-1} h \int_{t_l}^{t_{l+1}} \left(\int_{t_l}^s E \left[\left\| e^{A(s-u)} B(X_u) \right\|_{L_{HS}(U_0, H)}^2 \right] du \right) ds \right)^{\frac{1}{2}}.
\end{aligned}$$

Due to (A1)-(A4) again we finally obtain

$$\left(E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (F(X_s) - F(X_{t_l})) ds \right\|_H^2 \right] \right)^{\frac{1}{2}}$$

$$\begin{aligned}
&\leq C_T h^\gamma + C \sum_{l=0}^{m-1} \left(h \int_{t_l}^{t_{l+1}} (s - t_l)^2 ds \right)^{\frac{1}{2}} \\
&\quad + \left(C \sum_{l=0}^{m-1} h \int_{t_l}^{t_{l+1}} \int_{t_l}^s E \left[\|(-A)^{-\delta}\|_{L(H)}^2 \|B(X_u)\|_{L_{HS}(U_0, H_\delta)}^2 \right] du ds \right)^{\frac{1}{2}} \\
&\leq C_T h^\gamma + CMh^2 + \left(C \sum_{l=0}^{m-1} h \int_{t_l}^{t_{l+1}} (s - t_l) ds \right)^{\frac{1}{2}} \\
&\leq C_T h^\gamma + CMh^2 + C(Mh^3)^{\frac{1}{2}} \leq C_T h^\gamma.
\end{aligned}$$

The estimates of the second and third part follow easily by the triangular inequality, Hölders inequality and (A1)-(A4) as well

$$\begin{aligned}
&\left(E \left[\left\| \sum_{l=0}^{m-2} \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} - e^{A(t_m-t_l)} \right) F(X_{t_l}) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
&\leq \sum_{l=0}^{m-2} E \left[\left\| \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} - e^{A(t_m-t_l)} \right) F(X_{t_l}) ds \right\|_H^2 \right]^{\frac{1}{2}} \\
&\leq C \sum_{l=0}^{m-2} \left(h \int_{t_l}^{t_{l+1}} \|(-A)e^{A(t_m-s)}\|_{L(H)}^2 \|(-A)^{-1} (I - e^{A(s-t_l)})\|_{L(H)}^2 ds \right)^{\frac{1}{2}} \\
&\leq C \sum_{l=0}^{m-2} \left(h \int_{t_l}^{t_{l+1}} \left(\frac{s - t_l}{t_m - s} \right)^2 ds \right)^{\frac{1}{2}} \leq C \sum_{l=0}^{m-2} \left(h \int_{t_l}^{t_{l+1}} \left(\frac{s - t_l}{(m-l-1)h} \right)^2 ds \right)^{\frac{1}{2}} \\
&= C \sum_{l=0}^{m-2} \left(\frac{h^4}{(m-l-1)^2 h^2} \right)^{\frac{1}{2}} = Ch \sum_{l=0}^{m-2} \frac{1}{m-l-1} = Ch \sum_{l=1}^{m-1} \frac{1}{l} \leq C \frac{1 + \log M}{M} \\
&\leq C \frac{M^{1-\gamma}}{M(1-\gamma)} = CM^{-\gamma}.
\end{aligned}$$

Where in the last steps we employed some basic computations for $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$, see [18], $\sum_{l=1}^{m-1} \frac{1}{l} = 1 + \sum_{l=2}^{m-1} \frac{1}{l} \leq 1 + \sum_{l=2}^M \frac{1}{l} \leq 1 + \int_1^M \frac{1}{l} dl = 1 + \log M$ and for $r \in (0, 1]$ we have $1 + \ln(x) = 1 + \int_1^x s^{-1} ds \leq 1 + \int_1^x \frac{1}{s^{1-r}} ds = 1 + \frac{x^r - 1}{r} = \frac{x^r}{r} - \frac{(1-r)}{r} \leq \frac{x^r}{r}$.

Further we obtain

$$\begin{aligned}
&\left(E \left[\left\| \int_{t_{m-1}}^{t_m} \left(e^{A(t_m-s)} - e^{A(t_m-t_{m-1})} \right) F(X_{t_{m-1}}) ds \right\|_H^2 \right] \right)^{\frac{1}{2}} \\
&\leq \sqrt{h} \left(\int_{t_{m-1}}^{t_m} E \left[\left\| \left(e^{A(t_m-s)} - e^{A(t_m-t_{m-1})} \right) F(X_{t_{m-1}}) \right\|_H^2 \right] ds \right)^{\frac{1}{2}} \\
&\leq \sqrt{h} \left(\int_{t_{m-1}}^{t_m} C ds \right)^{\frac{1}{2}} \leq C_T M^{-1}.
\end{aligned}$$

5.3 Temporal discretization with Milstein scheme - diffusion B

For the estimation of the temporal error from the discretization of the stochastic integral we compute

$$E \left[\left\| \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} B(X_s) - e^{A(t_m-t_l)} B(X_{t_l}) \right) dW_s^K \right\|^2 \right]$$

$$\begin{aligned}
& - \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) dW_s^K \Bigg\|_H^2 \Bigg] \\
& \leq C \sum_{l=0}^{m-1} E \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (B(X_s) - B(X_{t_l})) dW_s^K \right. \right. \\
& \quad \left. \left. - \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) dW_s^K \right\|_H^2 \right] \\
& \quad + CE \left[\left\| \sum_{l=0}^{m-2} \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} - e^{A(t_m-t_l)} \right) B(X_{t_l}) dW_s^K \right\|_H^2 \right] \\
& \quad + CE \left[\left\| \int_{t_{m-1}}^{t_m} \left(e^{A(t_m-s)} - e^{A(t_m-t_{m-1})} \right) B(X_{t_{m-1}}) dW_s^K \right\|_H^2 \right] \\
& \leq C_T \left(M^{-\min(4(\gamma-\beta), 2\gamma)} + \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} \right)
\end{aligned}$$

where

$$\begin{aligned}
& \sum_{l=0}^{m-1} E \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (B(X_s) - B(X_{t_l})) dW_s^K \right. \right. \\
& \quad \left. \left. - \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) dW_s^K \right\|_H^2 \right] \\
& = \sum_{l=0}^{m-1} E \left[\left\| \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} \left(B'(X_{t_l})(X_s - X_{t_l}) \right. \right. \right. \right. \\
& \quad \left. \left. \left. + e^{A(t_m-s)} \int_0^1 B''(X_{t_l} + r(X_s - X_{t_l})) \left(X_s - X_{t_l}, X_s - X_{t_l} \right) dr \right) dW_s^K \right. \right. \\
& \quad \left. \left. - \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) dW_s^K \right\|_H^2 \right] \\
& = \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-s)} B'(X_{t_l}) \left((X_s - X_{t_l}) - \int_{t_l}^s B(X_{t_l}) dW_r^K \right) \right. \right. \\
& \quad \left. \left. + e^{A(t_m-s)} \int_0^1 B''(X_{t_l} + r(X_s - X_{t_l})) (X_s - X_{t_l}, X_s - X_{t_l}) dr \right\|_{L_{HS}(U_0, H)}^2 \right] ds
\end{aligned}$$

due to Itô's isometry. With Proposition 2.1 and Lemma 5.1 we obtain

$$\begin{aligned}
& \sum_{l=0}^{m-1} E \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (B(X_s) - B(X_{t_l})) dW_s^K \right. \right. \\
& \quad \left. \left. - \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) dW_s^K \right\|_H^2 \right] \\
& \leq 2 \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-s)} B'(X_{t_l}) \left((X_s - X_{t_l}) - \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) \right) \right\|_{L_{HS}(U_0, H)}^2 \right] ds \\
& \quad + 2 \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-s)} \right\|_{L(H)}^2 \int_0^1 \|B''(X_{t_l} + r(X_s - X_{t_l}))\|_{L^{(2)}(H_\beta, L_{HS}(U_0, H))}^2 \|X_s - X_{t_l}\|_{H_\beta}^4 \right] dr ds \\
& \leq C \sum_{l=0}^{m-1} \left(\int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-s)} B'(X_{t_l}) \left((X_s - X_{t_l}) - \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) \right) \right\|_{L_{HS}(U_0, H)}^2 \right] ds \right)
\end{aligned}$$

$$+ \frac{h^{1+\min(4(\gamma-\beta),2)}}{1+\min(4(\gamma-\beta),2)} \Big).$$

Next we plug in the expression for the mild solution

$$\begin{aligned} & \sum_{l=0}^{m-1} E \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (B(X_s) - B(X_{t_l})) \, dW_s^K \right. \right. \\ & \quad \left. \left. - \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) \, dW_r^K \right) \, dW_s^K \right\|_H^2 \right] \\ & \leq C \sum_{l=0}^{m-1} \left(\int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-s)} B'(X_{t_l}) \left((e^{A(s-t_l)} - I) X_{t_l} + \int_{t_l}^s e^{A(s-u)} F(X_u) \, du \right. \right. \right. \right. \\ & \quad \left. \left. + \int_{t_l}^s e^{A(s-u)} B(X_u) (dW_u - dW_u^K) + \int_{t_l}^s (e^{A(s-u)} - I) B(X_u) \, dW_u^K \right. \right. \\ & \quad \left. \left. + \int_{t_l}^s (B(X_u) - B(X_{t_l})) \, dW_u^K \right\|_{LHS(U_0, H)}^2 \right] \, ds + h^{1+\min(4(\gamma-\beta),2)} \Big) \\ & \leq C \sum_{l=0}^{m-1} \left(\int_{t_l}^{t_{l+1}} E \left[\left\| (e^{A(s-t_l)} - I) X_{t_l} \right\|_H^2 \right] \, ds + \int_{t_l}^{t_{l+1}} E \left[\left\| \int_{t_l}^s e^{A(s-u)} F(X_u) \, du \right\|_H^2 \right] \, ds \right. \\ & \quad \left. + \underbrace{\int_{t_l}^{t_{l+1}} E \left[\left\| \int_{t_l}^s e^{A(s-u)} B(X_u) (dW_u - dW_u^K) \right\|_H^2 \right] \, ds}_{\leq C_T h \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha}} \right. \\ & \quad \left. + \int_{t_l}^{t_{l+1}} E \left[\left\| \int_{t_l}^s (e^{A(s-u)} - I) B(X_u) \, dW_u^K \right\|_H^2 \right] \, ds \right. \\ & \quad \left. + \int_{t_l}^{t_{l+1}} E \left[\left\| \int_{t_l}^s (B(X_u) - B(X_{t_l})) \, dW_u^K \right\|_H^2 \right] \, ds + h^{1+\min(4(\gamma-\beta),2)} \Big). \end{aligned}$$

The proof of

$$\int_{t_l}^{t_{l+1}} E \left[\left\| \int_{t_l}^s e^{A(s-u)} B(X_u) (dW_u - dW_u^K) \right\|_H^2 \right] \, ds \leq C_T h \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha}$$

can be found in Section 5.4.

With Lemma 5.1 again and (A1)-(A4) this can further be rewritten as

$$\begin{aligned} & \sum_{l=0}^{m-1} E \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} (B(X_s) - B(X_{t_l})) \, dW_s^K \right. \right. \\ & \quad \left. \left. - \int_{t_l}^{t_{l+1}} e^{A(t_m-s)} B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) \, dW_r^K \right) \, dW_s^K \right\|_H^2 \right] \\ & \leq \sum_{l=0}^{m-1} \left(\int_{t_l}^{t_{l+1}} \left\| (-A)^{-\gamma} (e^{A(s-t_l)} - I) \right\|_{L(H)}^2 E \left[\| (-A)^\gamma X_{t_l} \|_H^2 \right] \, ds \right. \\ & \quad \left. + \int_{t_l}^{t_{l+1}} (s - t_l) \left(\int_{t_l}^s E \left[\left\| e^{A(s-u)} F(X_u) \right\|_H^2 \right] \, du \right) \, ds + C_T h \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} \right) \end{aligned}$$

$$\begin{aligned}
& + \int_{t_l}^{t_{l+1}} \int_{t_l}^s \left\| (-A)^{-\delta} \left(e^{A(s-u)} - I \right) \right\|_{L(H)}^2 E \left[\left\| (-A)^\delta B(X_u) \right\|_{L_{HS}(U_0, H)}^2 \right] du ds \\
& + \int_{t_l}^{t_{l+1}} \left(\int_{t_l}^s E \left[\left\| (B(X_u) - B(X_{t_l})) \right\|_{L_{HS}(U_0, H)}^2 \right] du \right) ds + h^{1+\min(4(\gamma-\beta), 2)} \\
& \leq C \sum_{l=0}^{m-1} \left(h^{2\gamma+1} + h^3 + C_T h \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} + h^{2\delta+2} + \int_{t_l}^{t_{l+1}} \left(\int_{t_l}^s (u - t_l)^{\min(2(\gamma-\beta), 1)} du \right) ds \right. \\
& \quad \left. + h^{1+\min(4(\gamma-\beta), 2)} \right) \\
& \leq C \sum_{l=0}^{m-1} \left(h^{2\gamma+1} + h^3 + h^{2\delta+2} + C_T h \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} + h^{\min(2(\gamma-\beta), 1)+2} + h^{1+\min(4(\gamma-\beta), 2)} \right) \\
& \leq C_T \left(C_T \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} + h^{\min(4(\gamma-\beta), 2\gamma)} \right)
\end{aligned}$$

where we used $2 + 2(\gamma - \beta) \geq 1 + 4(\gamma - \beta)$.

The second term is estimated for all $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$, using the independence of the increments of the Q -Wiener process in time, the Itô isometry and (A1)-(A4)

$$\begin{aligned}
& E \left[\left\| \sum_{l=0}^{m-2} \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} - e^{A(t_m-t_l)} \right) B(X_{t_l}) dW_s^K \right\|_H^2 \right] \\
& = \sum_{l=0}^{m-2} E \left[\left\| \int_{t_l}^{t_{l+1}} \left(e^{A(t_m-s)} - e^{A(t_m-t_l)} \right) B(X_{t_l}) dW_s^K \right\|_H^2 \right] \\
& \leq \sum_{l=0}^{m-2} \int_{t_l}^{t_{l+1}} \left\| (-A)^{-\delta} \left(e^{A(t_m-s)} - e^{A(t_m-t_l)} \right) \right\|_{L(H)}^2 E \left[\left\| (-A)^\delta B(X_{t_l}) \right\|_{L_{HS}(U_0, H)}^2 \right] ds \\
& \leq C \sum_{l=0}^{m-2} \int_{t_l}^{t_{l+1}} \left\| (-A)^{1-\delta} e^{A(t_m-s)} \right\|_{L(H)}^2 \underbrace{\left\| (-A)^{-1} \left(I - e^{A(s-t_l)} \right) \right\|_{L(H)}^2}_{\leq (s-t_l)^2 \leq h^2} ds \\
& \leq Ch^2 \sum_{l=0}^{m-2} \int_{t_l}^{t_{l+1}} (t_m - s)^{2(\delta-1)} ds = Ch^2 \sum_{l=0}^{m-2} \left((t_m - t_{l+1})^{2\delta-1} - (t_m - t_l)^{2\delta-1} \right) \\
& = Ch^2 \left((t_m - t_{m-1})^{2\delta-1} - (t_m)^{2\delta-1} \right) \leq C_T h^{2\delta+1} \leq C_T h^{2\gamma}.
\end{aligned}$$

Finally we obtain by (A1)-(A4)

$$\begin{aligned}
& E \left[\left\| \int_{t_{m-1}}^{t_m} \left(e^{A(t_m-s)} - e^{A(t_m-t_{m-1})} \right) B(X_{t_{m-1}}) dW_s^K \right\|_H^2 \right] \\
& \leq C \int_{t_{m-1}}^{t_m} \left\| e^{A(t_m-s)} \right\|_{L(H)}^2 \left\| (-A)^{-\delta} \left(I - e^{A(s-t_{m-1})} \right) \right\|_{L(H)}^2 E \left[\left\| (-A)^\delta B(X_{t_{m-1}}) \right\|_{L_{HS}(U_0, H)}^2 \right] ds \\
& \leq Ch^{2\delta+1} \leq Ch^{2\gamma}.
\end{aligned}$$

5.4 Approximation of the Q -Wiener process

Next we prove the error estimate resulting from the approximation of the Q -Wiener process and employ

$$dW_s - dW_s^K = \sum_{\substack{j \in \mathcal{J} \\ \eta_j \neq 0}} \sqrt{\eta_j} e_j d\beta_s^j - \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \sqrt{\eta_j} e_j d\beta_s^j = \sum_{\substack{j \in \mathcal{J} \setminus \mathcal{J}_K \\ \eta_j \neq 0}} \sqrt{\eta_j} e_j d\beta_s^j.$$

for $s \in [0, T]$, $K \in \mathbb{N}$. For all $s \in [0, T]$, $l \in \{0, \dots, M\}$, $M \in \mathbb{N}$,

$$\begin{aligned}
& E \left[\left\| \int_{t_l}^s e^{A(s-u)} B(X_u) (dW_u - dW_u^K) \right\|_H^2 \right]^{\frac{1}{2}} \\
&= E \left[\left\| \sum_{\substack{j \in \mathcal{J} \setminus \mathcal{J}_K \\ \eta_j \neq 0}} \int_{t_l}^s e^{A(s-u)} B(X_u) \sqrt{\eta_j} d\beta_j \tilde{e}_j \right\|_H^2 \right]^{\frac{1}{2}} \\
&= \left(\sum_{\substack{j \in \mathcal{J} \setminus \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j \int_{t_l}^s E \left[\left\| e^{A(s-u)} B(X_u) Q^{-\alpha} Q^\alpha \tilde{e}_j \right\|_H^2 \right] du \right)^{\frac{1}{2}} \\
&= \left(\sum_{\substack{j \in \mathcal{J} \setminus \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j^{2\alpha+1} \int_{t_l}^s E \left[\left\| e^{A(s-u)} B(X_u) Q^{-\alpha} \tilde{e}_j \right\|_H^2 \right] du \right)^{\frac{1}{2}} \\
&\leq \left(\left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} \int_{t_l}^s E \left[\sum_{\substack{j \in \mathcal{J} \\ \eta_j \neq 0}} \eta_j \left\| e^{A(s-u)} B(X_u) Q^{-\alpha} \tilde{e}_j \right\|_H^2 \right] du \right)^{\frac{1}{2}} \\
&= \left(\left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} \int_{t_l}^s E \left[\left\| e^{A(s-u)} B(X_u) Q^{-\alpha} \right\|_{L_{HS}(U_0, H)}^2 \right] du \right)^{\frac{1}{2}}.
\end{aligned}$$

By assumption (A1)-(A4) we get

$$\begin{aligned}
& E \left[\left\| \int_{t_l}^s e^{A(s-u)} B(X_u) (dW_u - dW_u^K) \right\|_H^2 \right]^{\frac{1}{2}} \\
&\leq \left(\left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} \int_{t_l}^s \|(-A)^\vartheta e^{A(s-u)}\|_{L(H)}^2 E \left[\left\| (-A)^{-\vartheta} B(X_u) Q^{-\alpha} \right\|_{L_{HS}(U_0, H)}^2 \right] du \right)^{\frac{1}{2}} \\
&\leq \left(C \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} \int_{t_l}^s (s-u)^{-2\vartheta} du \right)^{\frac{1}{2}} = \left(C \left(\sup_{j \in \mathcal{J} \setminus \mathcal{J}_K} \eta_j \right)^{2\alpha} \frac{(t_l-s)^{-2\vartheta+1}}{-2\vartheta+1} \right)^{\frac{1}{2}}.
\end{aligned}$$

5.5 Proof of (18)

Finally for $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$, we estimate

$$\begin{aligned}
E \left[\left\| \bar{X}_{t_m} - Y_m \right\|_H^2 \right] &= E \left[\left\| P_N \left(\sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} (F(X_{t_l}) - F(Y_l)) ds \right. \right. \right. \\
&\quad \left. \left. + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} (B(X_{t_l}) - B(Y_l)) dW_s^K \right. \right. \\
&\quad \left. \left. + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} \left(B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) \right. \right. \right. \\
&\quad \left. \left. \left. - B'(Y_l) \left(\int_{t_l}^s B(Y_l) dW_r^K \right) \right) dW_s^K \right\|_H^2 \right] \\
&\leq 3 \left(Mh \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-t_l)} (F(X_{t_l}) - F(Y_l)) \right\|_H^2 \right] ds \right. \\
&\quad \left. + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-t_l)} (B(X_{t_l}) - B(Y_l)) \right\|_{L_{HS}(U_0, H)}^2 \right] ds \right)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-t_l)} \left(B'(X_{t_l}) \left(\int_{t_l}^s B(X_{t_l}) dW_r^K \right) \right. \right. \right. \\
& \quad \left. \left. \left. - B'(Y_l) \left(\int_{t_l}^s B(Y_l) dW_r^K \right) \right) \right\|_{L_{HS}(U_0, H)}^2 \right] ds \\
& \leq C_T h \sum_{l=0}^{m-1} E \left[\|F(X_{t_l}) - F(Y_l)\|_H^2 \right] + Ch \sum_{l=0}^{m-1} E \left[\|B(X_{t_l}) - B(Y_l)\|_{L_{HS}(U_0, H)}^2 \right] \\
& \quad + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| e^{A(t_m-t_l)} \left(B'(X_{t_l}) \left(\sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \int_{t_l}^s B(X_{t_l}) \tilde{e}_j \sqrt{\eta_j} d\beta_r^j \right) \right. \right. \right. \\
& \quad \left. \left. \left. - B'(Y_l) \left(\sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \int_{t_l}^s B(Y_l) \tilde{e}_j \sqrt{\eta_j} d\beta_r^j \right) \right) \right\|_{L_{HS}(U_0, H)}^2 \right] ds.
\end{aligned}$$

By assumptions (A2), (A3) and the properties of the independent Brownian motions $(\beta_t^j)_{t \in [0, T], j \in \mathcal{J}}$ we obtain

$$\begin{aligned}
E \left[\|\bar{X}_{t_m} - Y_m\|_H^2 \right] & \leq C_T h \sum_{l=0}^{m-1} E \left[\|X_{t_l} - Y_l\|_H^2 \right] + Ch \sum_{l=0}^{m-1} E \left[\|X_{t_l} - Y_l\|_H^2 \right] \\
& \quad + C \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| \left(B'(X_{t_l}) \left(\sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} B(X_{t_l}) \tilde{e}_j \sqrt{\eta_j} (\beta_s^j - \beta_{t_l}^j) \right) \right. \right. \right. \\
& \quad \left. \left. \left. - B'(Y_l) \left(\sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} B(Y_l) \tilde{e}_j \sqrt{\eta_j} (\beta_s^j - \beta_{t_l}^j) \right) \right) \right\|_{L_{HS}(U_0, H)}^2 \right] ds
\end{aligned}$$

and

$$\begin{aligned}
& E \left[\|\bar{X}_{t_m} - Y_m\|_H^2 \right] \\
& \leq C_T h \sum_{l=0}^{m-1} E \left[\|X_{t_l} - Y_l\|_H^2 \right] \\
& \quad + C \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \sqrt{\eta_j} (B'(X_{t_l}) (B(X_{t_l}) \tilde{e}_j) - B'(Y_l) (B(Y_l) \tilde{e}_j)) (\beta_s^j - \beta_{t_l}^j) \right\|_{L_{HS}(U_0, H)}^2 \right] ds \\
& \leq C_T h \sum_{l=0}^{m-1} E \left[\|X_{t_l} - Y_l\|_H^2 \right] \\
& \quad + C \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} \sum_{\substack{j \in \mathcal{J} \\ \eta_j \neq 0}} \eta_j E \left[\left\| (B'(X_{t_l}) (B(X_{t_l}) \tilde{e}_j) - B'(Y_l) (B(Y_l) \tilde{e}_j)) \right\|_{L_{HS}(U_0, H)}^2 \right] E \left[(\beta_s^j - \beta_{t_l}^j)^2 \right] ds \\
& \leq C_T h \sum_{l=0}^{m-1} E \left[\|X_{t_l} - Y_l\|_H^2 \right] \\
& \quad + C \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| B'(X_{t_l}) (B(X_{t_l})) - B'(Y_l) (B(Y_l)) \right\|_{L_{HS}^{(2)}(U_0, H)}^2 \right] (s - t_l) ds \\
& \leq C_T h \sum_{l=0}^{m-1} E \left[\|X_{t_l} - Y_l\|_H^2 \right].
\end{aligned}$$

5.6 Approximation of the derivative

It remains to show that the approximation of the derivative does not distort the convergence properties. In the proof we need the following lemma

Lemma 5.2. *Under assumptions (A1)-(A4) it holds for all $p \geq 2$, $M \in \mathbb{N}$,*

$$\sup_{m \in \{0, \dots, M\}} E \left[\|Y_m\|_{H_\delta}^p \right]^{\frac{1}{p}} \leq C_{p,T,Q} \left(1 + E \left[\|\xi\|_{H_\delta}^p \right]^{\frac{1}{p}} \right).$$

Proof of Lemma 5.2. We get by the triangle inequality

$$\begin{aligned} E[\|Y_m\|_{H_\delta}^p]^{\frac{2}{p}} &\leq \left(CE \left[\|X_0\|_{H_\delta}^p \right]^{\frac{1}{p}} + \sum_{l=0}^{m-1} E \left[\left\| \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} F(Y_l) ds \right\|_{H_\delta}^p \right]^{\frac{1}{p}} \right. \\ &\quad + E \left[\left\| \int_{t_0}^{t_m} \sum_{l=0}^{m-1} e^{A(t_m-t_l)} B(Y_l) \mathbf{1}_{[t_l, t_{l+1})}(s) dW_s^K \right\|_{H_\delta}^p \right]^{\frac{1}{p}} \\ &\quad + \sum_{l=0}^{m-1} E \left[\left\| e^{A(t_m-t_l)} \frac{1}{\sqrt{h}} \left(B \left(Y_l + \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right) - B(Y_l) \right) \Delta W_l^K \right\|_{H_\delta}^p \right]^{\frac{1}{p}} \\ &\quad \left. + \sum_{l=0}^{m-1} E \left[\left\| e^{A(t_m-t_l)} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \bar{B}(Y_l, h, j) \sqrt{\eta_j} \tilde{e}_j \right\|_{H_\delta}^p \right]^{\frac{1}{p}} \right)^2. \end{aligned}$$

By Hölders inequality and Lemma 7.7 in [7] we obtain for $p \geq 2$

$$\begin{aligned} E[\|Y_m\|_{H_\delta}^p]^{\frac{2}{p}} &\leq C \left(E \left[\|X_0\|_{H_\delta}^p \right]^{\frac{2}{p}} + \left(\sum_{l=0}^{m-1} \left(\int_{t_l}^{t_{l+1}} E \left[\left\| (-A)^\delta e^{A(t_m-t_l)} F(Y_l) \right\|_H^p \right] ds \right)^{\frac{1}{p}} h^{1-\frac{1}{p}} \right)^2 \right. \\ &\quad + \int_{t_0}^{t_m} E \left[\left\| \sum_{l=0}^{m-1} e^{A(t_m-t_l)} B(Y_l) \mathbf{1}_{[t_l, t_{l+1})}(s) \right\|_{L_{HS}(U_0, H_\delta)}^p \right]^{\frac{2}{p}} ds \\ &\quad + \left(\sum_{l=0}^{m-1} \|(-A)^\delta e^{A(t_m-t_l)}\|_{L(H)} E \left[\left\| \frac{1}{\sqrt{h}} \left(B \left(Y_l + \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right) - B(Y_l) \right) \Delta W_l^K \right\|_H^p \right]^{\frac{1}{p}} \right)^2 \\ &\quad \left. + \left(\sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} E \left[\left\| (-A)^\delta e^{A(t_m-t_l)} \bar{B}(Y_l, h, j) \sqrt{\eta_j} \tilde{e}_j \right\|_H^p \right]^{\frac{1}{p}} \right)^2 \right). \end{aligned}$$

We now concentrate on

$$\bar{B}(Y_l, h, j) = B \left(Y_l - \frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right) - B(Y_l)$$

and use the following Taylor expansions of the difference approximations

$$B \left(Y_l + \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right) \Delta W_l^K = B(Y_l) \Delta W_l^K + B'(\xi(Y_l, \Delta W_l^K)) \left(\frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K, \Delta W_l^K \right) \quad (19)$$

$$B \left(Y_l - \frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right) \sqrt{\eta_j} \tilde{e}_j = B(Y_l) \sqrt{\eta_j} \tilde{e}_j + B'(\xi(Y_l, j)) \left(-\frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j, \sqrt{\eta_j} \tilde{e}_j \right). \quad (20)$$

Due to (A1) - (A3), Lemma 5.1 and Lemma 7.7 in [7] we get

$$\begin{aligned}
E[\|Y_m\|_{H_\delta}^p]^{\frac{2}{p}} &\leq CE \left[\|X_0\|_{H_\delta}^p \right]^{\frac{2}{p}} + Ch^{2-\frac{2}{p}} M \sum_{l=0}^{m-1} \left(\int_{t_l}^{t_{l+1}} (t_m - t_l)^{-\delta p} ds \right)^{\frac{2}{p}} E \left[\|F(Y_l)\|_H^p \right]^{\frac{2}{p}} \\
&\quad + C \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} E \left[\left\| \sum_{l=0}^{m-1} e^{A(t_m-t_l)} B(Y_l) \mathbf{1}_{[t_l, t_{l+1})}(s) \right\|_{L_{HS}(U_0, H_\delta)}^p \right]^{\frac{2}{p}} ds \\
&\quad + C \frac{M}{h} \sum_{l=0}^{m-1} (t_m - t_l)^{-2\delta} E \left[\left\| B'(\xi(Y_l, \Delta W_l^K)) \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right\|_{L(U, H)}^p \|\Delta W_l^K\|_U^p \right]^{\frac{2}{p}} \\
&\quad + CM \sum_{l=0}^{m-1} \left(\sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} (t_m - t_l)^{-\delta} E \left[\left\| -B'(\xi(Y_l, j)) \frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right\|_{L(U, H)}^p \|\sqrt{\eta_j} \tilde{e}_j\|_U^p \right]^{\frac{1}{p}} \right)^2 \\
&\leq CE \left[\|X_0\|_{H_\delta}^p \right]^{\frac{2}{p}} + C_{p,T} h^{1-\frac{2}{p}} \sum_{l=0}^{m-1} \left(h(t_m - t_l)^{-\delta p} \right)^{\frac{2}{p}} (1 + E[\|Y_l\|_{H_\delta}^p]^{\frac{2}{p}}) \\
&\quad + C \sum_{l=0}^{m-1} E \left[\|B(Y_l)\|_{L_{HS}(U_0, H_\delta)}^p \right]^{\frac{2}{p}} \int_{t_l}^{t_{l+1}} \left\| (-A)^{-\delta} \right\|_{L(H)}^2 \left\| (-A)^\delta e^{A(t_m-t_l)} \right\|_{L(H)}^2 ds \\
&\quad + CM \sum_{l=0}^{m-1} (t_m - t_l)^{-2\delta} E \left[\|B'(\xi(Y_l, \Delta W_l^K))\|_{L(H, L(U, H))}^p \|B(Y_l) \Delta W_l^K\|_{H_\delta}^p \|\Delta W_l^K\|_U^p \right]^{\frac{2}{p}} \\
&\quad + CM \sum_{l=0}^{m-1} (t_m - t_l)^{-2\delta} \left(\sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \sqrt{\eta_j} h E \left[\|B'(\xi(Y_l, j))\|_{L(H, L(U, H))}^p \|B(Y_l) \sqrt{\eta_j} \tilde{e}_j\|_{H_\delta}^p \right]^{\frac{1}{p}} \right)^2 \\
&\leq CE \left[\|X_0\|_{H_\delta}^p \right]^{\frac{2}{p}} + h^{1-2\delta} C_{p,T} \sum_{l=0}^{m-1} (m-l)^{-2\delta} (1 + E[\|Y_l\|_{H_\delta}^p]^{\frac{2}{p}}) \\
&\quad + C_p \sum_{l=0}^{m-1} h(t_m - t_l)^{-2\delta} (1 + E[\|Y_l\|_{H_\delta}^p]^{\frac{2}{p}}) \\
&\quad + C_p M h^{-2\delta} \sum_{l=0}^{m-1} (m-l)^{-2\delta} E \left[\|Y_l\|_{H_\delta}^p \|\Delta W_l^K\|_U^{2p} \right]^{\frac{2}{p}} \\
&\quad + C_{p,T} h^{1-2\delta} \sum_{l=0}^{m-1} (m-l)^{-2\delta} \left(\sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \sqrt{\eta_j} E \left[\|Y_l\|_{H_\delta}^p \|\sqrt{\eta_j} \tilde{e}_j\|_U^p \right]^{\frac{1}{p}} \right)^2.
\end{aligned}$$

Similar as in [16] we obtain for $\delta \in (0, \frac{1}{2})$

$$\sum_{l=0}^{m-1} (m-l)^{-2\delta} = \sum_{l=1}^m \frac{1}{l^{2\delta}} \leq 1 + \int_1^M \frac{1}{r^{2\delta}} dr = 1 + \frac{M^{1-2\delta} - 1}{1-2\delta} \leq \frac{M^{1-2\delta}}{1-2\delta}. \quad (21)$$

Therefore we get

$$\begin{aligned}
E[\|Y_m\|_{H_\delta}^p]^{\frac{2}{p}} &\leq CE \left[\|X_0\|_{H_\delta}^p \right]^{\frac{2}{p}} + h^{1-2\delta} C_{T,p,Q} \sum_{l=0}^{m-1} (m-l)^{-2\delta} (1 + E[\|Y_l\|_{H_\delta}^p]^{\frac{2}{p}}) \\
&\leq CE \left[\|X_0\|_{H_\delta}^p \right]^{\frac{2}{p}} + C_{T,p,Q} + h^{1-2\delta} C_{T,p,Q} \sum_{l=0}^{m-1} (m-l)^{-2\delta} E[\|Y_l\|_{H_\delta}^p]^{\frac{2}{p}}.
\end{aligned}$$

By the discrete Gronwall lemma we finally obtain

$$E[\|Y_m\|_{H_\delta}^p]^{\frac{2}{p}} \leq \left(C_p E \left[\|X_0\|_{H_\delta}^p \right]^{\frac{2}{p}} + C_{T,p,Q} \right) e^{C_{T,p,Q} h^{1-2\delta} \sum_{l=0}^{m-1} (m-l)^{-2\delta}}$$

$$\leq C_{T,p,Q} \left(1 + E \left[\|X_0\|_{H_\delta}^p \right]^{\frac{2}{p}} \right).$$

For the DFMM scheme with $l \in \{0, \dots, M\}$, $M \in \mathbb{N}$, $l \in \mathcal{J}$, and

$$\bar{B}(Y_l, h, j) = \left(b \left(\cdot, Y_l - \frac{h}{2} P_N b(\cdot, Y_l) \right) - b(\cdot, Y_l) \right) \eta_j \tilde{e}_j$$

the result follows analogously. \square

With this at hand we can proof the last estimate in (17) for all $m \in \{0, \dots, M\}$, $M \in \mathbb{N}$,

$$\begin{aligned} & E \left[\|\bar{Y}_{t_m} - Y_m\|_H^2 \right] \\ = & E \left[\left\| P_N \left(e^{A t_m} X_0 + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} F(Y_l) ds + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} B(Y_l) dW_s^K \right. \right. \right. \\ & \left. \left. + \sum_{l=0}^{m-1} \left(\frac{1}{2} e^{A(t_m-t_l)} B'(Y_l) (B(Y_l) \Delta W_l^K, \Delta W_l^K) - \frac{h}{2} e^{A(t_m-t_l)} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j B'(Y_l) (B(Y_l) \tilde{e}_j, \tilde{e}_j) \right) \right) \right. \\ & \left. - P_N \left(e^{A t_m} X_0 + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} F(Y_l) ds + \sum_{l=0}^{m-1} \int_{t_l}^{t_{l+1}} e^{A(t_m-t_l)} B(Y_l) dW_s^K \right. \right. \\ & \left. \left. + \sum_{l=0}^{m-1} e^{A(t_m-t_l)} \frac{1}{\sqrt{h}} \left(B \left(Y_l + \frac{1}{2} \sqrt{h} P_N B(Y_l) \Delta W_l^K \right) - B(Y_l) \right) \Delta W_l^K \right. \right. \\ & \left. \left. + \sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} e^{A(t_m-t_l)} \bar{B}(Y_l, h, j) \sqrt{\eta_j} \tilde{e}_j \right) \right\|_H^2 \right]. \end{aligned}$$

We are left with

$$\begin{aligned} & E \left[\|\bar{Y}_{t_m} - Y_m\|_H^2 \right] \\ \leq & E \left[\left\| P_N e^{A(t_m-t_l)} \left(\sum_{l=0}^{m-1} \left(\frac{1}{2} B'(Y_l) (B(Y_l) \Delta W_l^K, \Delta W_l^K) - \frac{h}{2} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j B'(Y_l) (B(Y_l) \tilde{e}_j, \tilde{e}_j) \right) \right) \right. \right. \\ & \left. - P_N \left(\sum_{l=0}^{m-1} e^{A(t_m-t_l)} \frac{1}{\sqrt{h}} \left(B \left(Y_l + \frac{1}{2} \sqrt{h} P_N B(Y_l) \Delta W_l^K \right) - B(Y_l) \right) \Delta W_l^K \right) \right. \\ & \left. - P_N \left(\sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} e^{A(t_m-t_l)} \bar{B}(Y_l, h, j) \sqrt{\eta_j} \tilde{e}_j \right) \right\|_H^2 \right]. \end{aligned}$$

Again, we consider

$$\bar{B}(Y_l, h, j) = B \left(Y_l - \frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right) - B(Y_l).$$

first and use Taylor approximations similar to (19).

Inserting these expressions yields

$$\begin{aligned} & E \left[\|\bar{Y}_{t_m} - Y_m\|_H^2 \right] \\ \leq & E \left[\left\| P_N \left(\sum_{l=0}^{m-1} e^{A(t_m-t_l)} \left(\frac{1}{2} B'(Y_l) (B(Y_l) \Delta W_l^K, \Delta W_l^K) - \frac{h}{2} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \eta_j B'(Y_l) (B(Y_l) \tilde{e}_j, \tilde{e}_j) \right) \right) \right. \right. \end{aligned}$$

$$\begin{aligned}
& -P_N \left(\sum_{l=0}^{m-1} e^{A(t_m-t_l)} \left(\frac{1}{\sqrt{h}} B'(Y_l) \left(\frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K, \Delta W_l^K \right) \right. \right. \\
& \quad \left. \left. + \frac{1}{2} B''(\xi(Y_l, \Delta W_l^K)) \left(\frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K, \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right) \Delta W_l^K \right) \right) \\
& -P_N \left(\sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} e^{A(t_m-t_l)} \left(B'(Y_l) \left(-\frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j, \sqrt{\eta_j} \tilde{e}_j \right) \right. \right. \\
& \quad \left. \left. + \frac{1}{2} B''(\xi(Y_l, j)) \left(-\frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j, -\frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right) \sqrt{\eta_j} \tilde{e}_j \right) \right) \Bigg\|_H^2 \Bigg].
\end{aligned}$$

Further we rewrite

$$\begin{aligned}
& \mathbb{E} \left[\|\bar{Y}_{t_m} - Y_m\|_H^2 \right] \\
& \leq E \left[\left\| \sum_{l=0}^{m-1} e^{A(t_m-t_l)} \frac{1}{\sqrt{h}} \frac{1}{2} B''(\xi(Y_l, \Delta W_l^K)) \left(\frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K, \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right) \Delta W_l^K \right. \right. \\
& \quad \left. \left. - \sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} e^{A(t_m-t_l)} \frac{1}{2} B''(\xi(Y_l, j)) \left(\frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j, \frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right) \sqrt{\eta_j} \tilde{e}_j \right\|_H^2 \right] \\
& \leq C \left(E \left[\left\| \sum_{l=0}^{m-1} e^{A(t_m-t_l)} \frac{1}{\sqrt{h}} B''(\xi(Y_l, \Delta W_l^K)) \left(\frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K, \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right) \Delta W_l^K \right\|_H^2 \right]^{\frac{1}{2}} \right)^2 \\
& \quad + C \left(E \left[\left\| \sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} e^{A(t_m-t_l)} \frac{1}{2} B''(\xi(Y_l, j)) \left(\frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j, \frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right) \sqrt{\eta_j} \tilde{e}_j \right\|_H^2 \right]^{\frac{1}{2}} \right)^2.
\end{aligned}$$

Assumptions (A1) and (A3) yield

$$\begin{aligned}
& \mathbb{E} \left[\|\bar{Y}_{t_m} - Y_m\|_H^2 \right] \\
& \leq \left(\sum_{l=0}^{m-1} \frac{C}{\sqrt{h}} E \left[\left\| e^{A(t_m-t_l)} B''(\xi(Y_l, \Delta W_l^K)) \left(\frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K, \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right) \Delta W_l^K \right\|_H^2 \right]^{\frac{1}{2}} \right)^2 \\
& \quad + C \left(\sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} E \left[\left\| e^{A(t_m-t_l)} \frac{1}{2} B''(\xi(Y_l, j)) \left(\frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j, \frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right) \sqrt{\eta_j} \tilde{e}_j \right\|_H^2 \right]^{\frac{1}{2}} \right)^2 \\
& \leq \left(C \sum_{l=0}^{m-1} \frac{1}{\sqrt{h}} E \left[\|B''(\xi(Y_l, \Delta W_l^K))\|_{L^{(2)}(H, L(U, H))}^2 \left\| \frac{\sqrt{h}}{2} P_N B(Y_l) \Delta W_l^K \right\|_H^4 \|\Delta W_l^K\|_U^2 \right]^{\frac{1}{2}} \right)^2 \\
& \quad + \left(C \sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} E \left[\|B''(\xi(Y_l, j))\|_{L^{(2)}(H, L(U, H))}^2 \left\| \frac{h}{2} P_N B(Y_l) \sqrt{\eta_j} \tilde{e}_j \right\|_H^4 \|\sqrt{\eta_j} \tilde{e}\|_U^2 \right]^{\frac{1}{2}} \right)^2.
\end{aligned}$$

With (A1)-(A4) again and the fact that Q is trace class we obtain

$$\begin{aligned}
\mathbb{E} \left[\|\bar{Y}_{t_m} - Y_m\|_H^2 \right] & \leq \left(C \sum_{l=0}^{m-1} \frac{\sqrt{h}}{4} E \left[\|B(Y_l)\|_{L(U, H_\delta)}^4 \|\Delta W_l^K\|_U^6 \right]^{\frac{1}{2}} \right)^2 \\
& \quad + \left(C \sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} \frac{h^2}{4} \eta_j^{\frac{3}{2}} E \left[\|B(Y_l)\|_{L(U, H_\delta)}^4 \right]^{\frac{1}{2}} \right)^2
\end{aligned}$$

$$\begin{aligned}
&\leq \left(C \sum_{l=0}^{m-1} \sqrt{h} \left(1 + E \left[\|Y_l\|_{H_\delta}^4 \right] \right)^{\frac{1}{2}} E \left[\|\Delta W_l^K\|_U^6 \right]^{\frac{1}{2}} \right)^2 \\
&\quad + \left(C \sum_{l=0}^{m-1} \sum_{\substack{j \in \mathcal{J}_K \\ \eta_j \neq 0}} h^2 \eta_j^{\frac{3}{2}} \left(1 + E \left[\|Y_l\|_{H_\delta}^4 \right] \right)^{\frac{1}{2}} \right)^2 \\
&\leq \left(C \sum_{l=0}^{m-1} h^2 \left(C(1 + E[\|Y_l\|_{H_\delta}^4]) \right)^{\frac{1}{2}} \right)^2 \\
&\quad + \left(C \sum_{l=0}^{m-1} \left(\sup_{j \in \mathcal{J}_K} \sqrt{\eta_j} \right) \text{tr} Q h^2 \left(1 + E[\|Y_l\|_{H_\delta}^4] \right)^{\frac{1}{2}} \right)^2 \\
&\leq \left(C \sum_{l=0}^{m-1} h^2 \right)^2 + 2 \left(C \sum_{l=0}^{m-1} \left(\sup_{j \in \mathcal{J}} \sqrt{\eta_j} \right) \text{tr} Q \frac{h^2}{4} \right)^2 \leq C_{T,Q} h^2.
\end{aligned}$$

This proves the error estimate for the general case.

Now let $l \in \{0, \dots, M\}$, $M \in \mathbb{N}$, $j \in \mathcal{J}$, for

$$\bar{B}(Y_l, h, j) = \left(b \left(\cdot, Y_l - \frac{h}{2} P_N b(\cdot, Y_l) \right) - b(\cdot, Y_l) \right) \eta_j \tilde{e}_j$$

we use

$$\begin{aligned}
b \left(\cdot, Y_l - \frac{h}{2} P_N b(\cdot, Y_l) \right) \eta_j \tilde{e}_j^2 &= b(\cdot, Y_l) \eta_j \tilde{e}_j^2 + b'(\cdot, Y_l) \cdot \left(-\frac{h}{2} P_N b(\cdot, Y_l) \right) \eta_j \tilde{e}_j^2 \\
&\quad + \frac{1}{2} b''(\cdot, \xi(Y_l, j)) \cdot \left(-\frac{h}{2} P_N b(\cdot, Y_l) \right) \cdot \left(-\frac{h}{2} P_N b(\cdot, Y_l) \right) \eta_j \tilde{e}_j^2
\end{aligned}$$

and the estimate

$$E \left[\|\bar{Y}_{t_m} - Y_m\|_H^2 \right] \leq C_{T,Q} h^2$$

follows as above. □

References

- [1] I. A. Adamu and G. J. Lord. Numerical approximation of multiplicative SPDEs. *Int. J. Comput. Math.*, 89(18):2603–2621, 2012.
- [2] A. Barth and A. Lang. Milstein approximation for advection-diffusion equations driven by multiplicative noncontinuous martingale noises. *Appl. Math. Optim.*, 66(3):387–413, 2012.
- [3] A. Barth and A. Lang. L^p and almost sure convergence of a Milstein scheme for stochastic partial differential equations. *Stochastic Process. Appl.*, 123(5):1563–1587, 2013.
- [4] S. Becker, A. Jentzen, and P. E. Kloeden. An exponential Wagner-Platen type scheme for SPDEs. *ArXiv e-prints*, Sept. 2013.
- [5] D. Blömker, M. Kamrani, and S. M. Hosseini. Full discretization of the stochastic Burgers equation with correlated noise. *IMA J. Numer. Anal.*, 33(3):825–848, 2013.
- [6] R. A. Carmona and M. R. Tehranchi. *Interest rate models: an infinite dimensional stochastic analysis perspective*. Springer Finance. Springer-Verlag, Berlin, 2006.

- [7] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.
- [8] A. Debussche. Weak approximation of stochastic partial differential equations: the nonlinear case. *Math. Comp.*, 80(273):89–117, 2011.
- [9] B. Ghayebi and S. M. Hosseini. A simplified Milstein scheme for SPDEs with multiplicative noise. *Abstr. Appl. Anal.*, pages Art. ID 140849, 15, 2014.
- [10] W. Grecksch and P. E. Kloeden. Time-discretised Galerkin approximations of parabolic stochastic PDEs. *Bull. Austral. Math. Soc.*, 54(1):79–85, 1996.
- [11] I. Gyöngy. Lattice approximations for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise. I. *Potential Anal.*, 9(1):1–25, 1998.
- [12] I. Gyöngy and D. Nualart. Implicit scheme for quasi-linear parabolic partial differential equations perturbed by space-time white noise. *Stochastic Process. Appl.*, 58(1):57–72, 1995.
- [13] E. Hausenblas. Approximation for semilinear stochastic evolution equations. *Potential Anal.*, 18(2):141–186, 2003.
- [14] E. Hausenblas. Finite element approximation of stochastic partial differential equations driven by Poisson random measures of jump type. *SIAM J. Numer. Anal.*, 46(1):437–471, 2007/08.
- [15] A. Jentzen, P. Kloeden, and G. Winkel. Efficient simulation of nonlinear parabolic SPDEs with additive noise. *Ann. Appl. Probab.*, 21(3):908–950, 2011.
- [16] A. Jentzen and P. E. Kloeden. Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 465(2102):649–667, 2009.
- [17] A. Jentzen and P. E. Kloeden. *Taylor approximations for stochastic partial differential equations*, volume 83 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.
- [18] A. Jentzen and M. Röckner. A Milstein scheme for SPDEs. *ArXiv e-prints*, Jan. 2012.
- [19] A. Jentzen and M. Röckner. Regularity analysis for stochastic partial differential equations with nonlinear multiplicative trace class noise. *J. Differential Equations*, 252(1):114–136, 2012.
- [20] A. Jentzen and M. Röckner. A Milstein scheme for SPDEs. *Found. Comput. Math.*, 15(2):313–362, 2015.
- [21] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
- [22] P. E. Kloeden and S. Shott. Linear-implicit strong schemes for Itô-Galerkin approximations of stochastic PDEs. *J. Appl. Math. Stochastic Anal.*, 14(1):47–53, 2001. Special issue: Advances in applied stochastics.
- [23] M. Kovács, S. Larsson, and F. Saedpanah. Finite element approximation of the linear stochastic wave equation with additive noise. *SIAM J. Numer. Anal.*, 48(2):408–427, 2010.
- [24] A. Lang, P. Chow, and J. Potthoff. Almost sure convergence for a semidiscrete Milstein scheme for SPDEs of Zakai type. *Stochastics*, 82(1-3):315–326, 2010.
- [25] A. Lang, P. Chow, and J. Potthoff. Erratum: Almost sure convergence of a semi-discrete Milstein scheme for SPDEs of Zakai type [mr2677551]. *Stochastics*, 84(4):561, 2012.
- [26] G. J. Lord and J. Rougemont. A numerical scheme for stochastic PDEs with Gevrey regularity. *IMA J. Numer. Anal.*, 24(4):587–604, 2004.

- [27] G. J. Lord and A. Tambue. Stochastic exponential integrators for the finite element discretization of SPDEs for multiplicative and additive noise. *IMA J. Numer. Anal.*, 33(2):515–543, 2013.
- [28] T. Müller-Gronbach and K. Ritter. Lower bounds and nonuniform time discretization for approximation of stochastic heat equations. *Found. Comput. Math.*, 7(2):135–181, 2007.
- [29] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.
- [30] C. Prévôt and M. Röckner. *A concise course on stochastic partial differential equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.
- [31] M. Renardy and R. C. Rogers. *An introduction to partial differential equations*, volume 13 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 2004.
- [32] A. Rößler. Second order Runge-Kutta methods for Stratonovich stochastic differential equations. *BIT*, 47(3):657–680, 2007.
- [33] A. Rößler. Second order Runge-Kutta methods for Itô stochastic differential equations. *SIAM J. Numer. Anal.*, 47(3):1713–1738, 2009.
- [34] A. Rößler. Runge-Kutta methods for the strong approximation of solutions of stochastic differential equations. *SIAM J. Numer. Anal.*, 48(3):922–952, 2010.
- [35] G. R. Sell and Y. You. *Dynamics of evolutionary equations*, volume 143 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2002.
- [36] H. C. Tuckwell. Stochastic partial differential equations in neurobiology: linear and nonlinear models for spiking neurons. In *Stochastic biomathematical models*, volume 2058 of *Lecture Notes in Math.*, pages 149–173. Springer, Heidelberg, 2013.
- [37] J. B. Walsh. Finite element methods for parabolic stochastic PDE’s. *Potential Anal.*, 23(1):1–43, 2005.
- [38] X. Wang and S. Gan. A Runge-Kutta type scheme for nonlinear stochastic partial differential equations with multiplicative trace class noise. *Numer. Algorithms*, 62(2):193–223, 2013.
- [39] A. G. Werschulz. *The computational complexity of differential and integral equations*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1991. An information-based approach, Oxford Science Publications.
- [40] Y. Yan. Galerkin finite element methods for stochastic parabolic partial differential equations. *SIAM J. Numer. Anal.*, 43(4):1363–1384 (electronic), 2005.